

Which way now for Europe?

Mr Nicholas Ridley's outburst last week, and his subsequent resignation from the British government, is based on an unsustainable theory of the genetics of national behaviour much like Hitler's causes.

THE surprise is not that Mr Nicholas Ridley, until the weekend Secretary of State for Trade and Industry, believes that British membership of the European Monetary Union would be a recipe for German economic domination, but that he put his guileless views to the editor of the *Spectator* in the knowledge that they would be reported. The simplest explanation is that Ridley is so used to talking in this vein with his political friends that he had persuaded himself that the publication of his views would cause no fuss.

It is nevertheless as well that Britain's European partners should know that at least some of those with whom they negotiate hold to the unreconstructed view that there is such a thing as national character, and that it is immutable over generations.

Even the British prime minister, Mrs Margaret Thatcher, is not immune from the fallacy underlying this supposition. In some ways, it is more damaging for the British government's reputation that it should have come to light, at the weekend, that the prime minister held a meeting in March with a group of academic historians at which a list of supposed German character defects was canvassed. The group eventually agreed that it had "no misgivings" about the present leadership of Germany, but wondered whether "some of the unhappy characteristics of the past" might reemerge in 10 or 20 years, with "just as destructive consequences".

This is a strange position for grown-up people to be taking, implying as it does a more rigid genetic determination of human behaviour than respectable biologists would think credible.

During the Ridley crisis, it has not been much remarked upon that this opinion has much in common with that of Adolf Hitler that Aryans (whoever they are) have desirable and genetically determined behaviour. It would, of course, be absurd that Britain's position in the negotiations that lie ahead over German unity and other European matters were to be based on a genetic theory for which there is no evidence, but which is both implausible and which could well have "destructive consequences" of its own. Luckily, the majority of the British cabinet is unlikely to be stampeded in that direction.

The Ridley/Hitler view is belied by the facts. How else is it possible to reconcile Germany's role as Europe's melting-pot for the past 45 years with the casual assumption of the Ridley school that it is a homogenous and

fearsome society? As it happens, the German characteristics that Ridley and his friends most fear are recognizably the products of social institutions created since the Second World War. An excellent secondary school system has engendered enviable competence, a politically independent Bundesbank has engendered financial stability and a high savings ratio and the division of legislative power between the federal and *Länder* governments, backed up by the constitutional court at Karlsruhe, have given Germany an explicit guarantee of personal legal rights of which people (not necessarily governments) elsewhere would be thoroughly envious if they knew more about the system. Germany, moreover, is a classless state, which Britain is not.

How can it help that the primitive roots of some British reluctance to join up with Europe should now have been publicly exposed? Because they are now explicit, and will attract the ridicule they deserve. That the British cabinet is divided on the issue is also a worthwhile piece of public information. The difficulty, for those with enlightened views in Britain as well as for those elsewhere in Europe, will be to distinguish clearly between unsustainable diffidence about the development of European institutions and that which has a legitimate basis. To what extent, for example, is British insistence that a united Germany should be a continuing member of the North Atlantic Treaty Organization (NATO) coloured by fears that the Third Reich will be recreated? And that NATO itself should then be transformed from an exclusively military organization into one of political and even economic dimensions when its continuing functions might more conveniently be subsumed into those of other institutions?

Luckily, the British politicians (ministers especially) who are known not to share the Ridley view of Europe will now be more free to be influential. In relation to the European Communities, for example, it will now be simpler for those who fear the protectionist tendencies of the European Communities to make their arguments explicitly, unclouded by suspicions that their hidden agenda is to keep Germany in its place. There is even reason to hope that the frank disagreement, based on philosophically different views of the origins of national characteristics, within the British government will now be argued out, and that those who hold to the unreconstructed Ridley view will be forced to concede defeat. That could be good for Europe as well as Britain. □



Farmers' GATT

The General Agreement on Tariffs and Trade, the basis of the world's prosperity, is now seriously threatened.

LAST week's summit meeting of the seven largest industrialized nations at Houston failed in its most important task — to find a formula that will allow a successful negotiation of the rules that govern international trade in agricultural produce. The issue is important because of the declared ambition of these same countries to negotiate an agreement on farm produce as a part of the attempt to extend the scope of the General Agreement on Tariffs and Trade, otherwise GATT. The negotiations have been under way for the past four years, but are due to end in December. The danger is that failure to agree on agriculture will bring the whole enterprise to a halt.

The consequences of that could be exceedingly serious. GATT is a club requiring of its members that they should follow certain rules in regulating their trade with other members. In particular, they are required in respect of manufactured goods not to use discriminatory legal and other devices for restricting imports of goods. It would, for example, be illegal under GATT for a country importing widgets of some kind to insist that imported widgets should come only from the manufacturers of neighbouring countries, or to say that it will allow the import of only a fixed number of them so that its domestic manufacturers could hope to build up their own businesses. Import tariffs are allowed, provided that they are not discriminatory, but even then there are rules that specify upper limits for substantial classes of manufactured goods. It is generally agreed that GATT has made possible the rapid growth of world trade since the Second World War, and that the whole world has profited as a consequence.

But GATT is not a once-and-for-all agreement, and can never be. For one thing, the diplomats of international trade cannot hope to anticipate all the changes of technique that lie ahead. For another, to the extent that the objective of free trade is to encourage between different countries an efficient division of labour of which Adam Smith (who died 200 years ago last week) would have approved, it is plain that the scope of the agreement must be steadily extended. In the current negotiations, for example, it has been agreed that the GATT agreement should be extended to cover financial services of various kinds. Countries which have been found to be less successful at manufacturing goods may nevertheless hope to pay for what they have to import by selling insurance, or other such services, to their trading partners.

Not before time, agricultural produce has now found its way onto the agenda. The explanation is simple — during the past decade, both the United States and the European Communities have offended trading partners such as Australia and Canada by their practice of subsidizing or otherwise encouraging exports of food. The European Communities, which spend \$10,000 million (out of a total

farm subsidy of nearly \$50,000 million) on export subsidies, are the worst offender. The trouble with last week's Houston meeting is that while the participants called for negotiations on this vexed issue, there is nothing in the declaration to ensure that even members of the European Communities, will negotiate in a different frame of mind than in the past.

The whole issue contains elements of the bizarre. The total cost of what Europeans call the Common Agricultural Policy is not simply the \$44,000 million the policy will have cost this year, but a larger subsidy of farmers by European taxpayers through the higher food prices people in Europe are compelled to pay for the privilege of growing their own food uneconomically. The same policy is likely to stand powerfully in the way of the fulfilment of Western Europe's ambitions to cement strengthened trading relationships with Eastern Europe — many states there will be unable to buy goods from the West because the levy on food imports into the Communities will prevent the customers paying in the obvious coinage — agricultural exports. Nobody would pretend that the Common Agricultural Policy could be swept away overnight without causing enormous disruption to rural communities in Western Europe, but not even the GATT negotiators are asking for that. The need, now, is merely to agree that the policy cannot last forever. Sadly, there are many members of the Communities who will not agree that even that stark truth should be stated openly.

In reality, GATT should be looking for much more. To the extent that GATT helps to determine trading relationships between developed and developing countries, it is generally restrictive. In the present negotiations, it has been agreed that a version of the Multi-Fibre Agreement, which allows rich countries to restrict imports of textiles from developing countries, should persist. The objective is, of course, to protect the textile industries of the rich countries. The consequences are that domestic consumers must pay extra for textiles, and that developing countries are less able to earn the funds with which to pay for the capital goods they need to buy from elsewhere. The consequences are that both the textile protectionists and their potential competitors are impoverished; the latter are saddled with low-tech occupations (or none at all) when they might be engaged in middle-tech occupations, the former by the extent to which the value they might be adding to raw material falls short of what they might be adding in more technically advanced pursuits.

If the present round of negotiations is eventually completed and signed as a treaty, the signatories will be painfully aware that they have signed an imperfect document — better than the existing treaty, but sufficiently far from the ideal as to be contemptible. That is why there is now a need to establish GATT on a different basis. The best solution would be that GATT should be turned into an international supranational court of mercantile law. The next best is that GATT members should be compelled to begin the next round of negotiations before the current round is complete. □

Hanford nuclear radiation doses assessed

■ "Serious" exposure for 13,000 residents

■ Thyroid disease study under way

Boston

MORE than 13,000 people living near the Hanford Nuclear Reservation in south-eastern Washington state received "significant" doses of radiation as a result of secret emissions from the facility in the 1940s, according to a study* released last week by an independent panel of researchers supported by the US Department of Energy. The study is the first comprehensive attempt to reconstruct the amount of radiation received by people living near Hanford, which is used to produce nuclear weapons, and calculates levels of exposure greater than those so far seen near any other nuclear facility in the United States.

The data released last week represent the completion of the first phase of a five-year, \$15-million study. The initial phase focused solely on the largest known releases of radioactivity from the facility: air emissions which occurred between 1944 and 1947 and discharges into the Columbia River between 1964 and 1966. From the air-borne incidents alone, according to the study, approximately five per cent of the 270,000 residents living in the vicinity of the plant during and after the Second World War may have accumulated doses of radiation in excess of 33 rads over a three-year period and a small number of infants and children could have accumulated doses of radiation to their thyroid glands as high as 2,900 rads in the same period. The current level of airborne radiation considered safe by the US government for civilians living near nuclear weapons plants is 0.025 rads per year and workers in nuclear power plants in the United States are limited to 5 rads exposure per year to their entire body.

James Watkins, Secretary of the Department of Energy, described the levels of exposure received by Hanford residents as "serious" in a press conference called before the report was released to the public. Members of the study's research panel emphasize that they have not attempted to predict the effects of radiation exposure on health. John Till, the panel's chair and president of a private radiological assessment firm, stated upon the report's release that the Phase I study "does not, nor was it designed to, link dose estimates with health effects. Our job," Till stressed, "is to estimate doses

people may have received." Phase I of the "dose reconstruction" study sought to find the most dramatic releases of radiation in the facility's 50-year history and to calculate the doses people may have received using a computer model that incorporates variables such as weather, geography, and a variety of 'pathways' through which radiation may have reached the population. Future phases of the project will seek to apply the model to a larger geographic area over a greater time span and to include an increasing number of radionuclides. Estimates of the residents' accumulated radiation exposure are thus likely to grow, at least to some extent, by the end of the fourth and final phase of the study in 1994.

At the same time, a separate epidemiological study is under way, sponsored by the Atlanta-based Centers for Disease Control, to determine the incidence of thyroid disease among the affected population. Till and other panel members said that they felt "strongly" that the radiation doses found so far justify research on health effects. Kenneth Kopecky, a biostatistician at the Fred Hutchinson Cancer Research Center in Seattle, Washington, who is affiliated with both studies, said that results from the CDC-funded 'Hanford Thyroid Disease Study' can be expected by 1993.

Both studies were begun after more than 19,000 pages of secret documents were released by the Department of Energy in 1986. The documents reveal that more than 400,000 curies of radioisotopes were released secretly into the atmosphere between 1944 and 1947.

The releases, far greater than any known to have occurred since, took place during the reprocessing of uranium, when radioactive fuel rods were dissolved in acid to extract plutonium for use in nuclear weapons. Although the process continued at Hanford for decades, changes in the technology used and in the filtration systems employed prevented further releases of similar magnitude.

Over the past two years, the panel led by Till has analysed the publicly released documents as well as another 20,000 pages of material, some of which are still classified. The study concluded that airborne releases of radioactive iodine 131 were likely to have been most dangerous for residents. Although iodine 131 decays to harmless levels within a few months, it would have been consumed by residents in milk from cows that grazed on contam-

inated grasses in the vicinity of Hanford.

The panel also estimated radiation doses received by residents who drank water and ate fish from the Columbia river, into which eight of Hanford's nine nuclear fuel production reactors discharged radioactive cooling water, and found them to be far lower than those received from airborne radioisotopes. The highest estimated exposure was 1.7 rads for phosphorus-32 accumulating in the gastrointestinal tract. But future phases of the study are expected to document higher levels of exposure as a more diverse set of radionuclides and a longer span of time are considered.

Predictably, the findings have caused uproar among Hanford residents. Some are calling for a public acknowledgement of wrongdoing from the Department of Energy and for medical compensation and assistance for those affected.

Although the project is sponsored by the Energy Department, its researchers were selected independently by the deans of four northwestern universities and its findings have yet to be officially sanctioned. A special disclaimer accompanying the report states that the information presented does not "necessarily reflect" the views of the US government or any agency thereof. Apart from Secretary Watkins preemptive comments, Department of Energy officials say that the agency will provide no immediate reaction to the findings.

Seth Shulman

BRITISH UNIVERSITIES

Council chief resigns

London

SIR Peter Swinnerton-Dyer is to resign as chief executive of the Universities Funding Council (UFC) at the end of March 1991. In a letter circulated to UFC staff, Swinnerton-Dyer, who is nearing retirement age, explains that he does not wish to move from London to Bristol, where the UFC will move its offices next year.

The resignation had been expected by some observers. Swinnerton-Dyer, chairman of the University Grants Committee until it evolved into the UFC in 1989, is seen as a less-than-enthusiastic supporter of some of the UFC's current plans. In line with government policy, the UFC intends to shift university funding away from the central provision of grants towards a system of fees. Lord Chilver, chairman of UFC, has said that the planned expansion of the university sector will be accompanied by a decline in the proportion of funds supplied through UFC. Peter Aldous



*Phase I of the Hanford Environmental Dose Reconstruction Project, July 1990, Pacific Northwest Laboratory, Richland, Washington 99352.

Troubles never come singly

Washington

CONTINUING problems in the US space programme this week prompted the White House to ask NASA (the US National Aeronautics and Space Administration) to set up a new task force of outside experts to consider the agency's long-term goals. The request came on Monday during a meeting between Vice-President Dan Quayle, chairman of the National Space Council, and NASA administrator Richard Truly. Earlier, it was rumoured that the council might force major changes upon NASA. But Quayle said a shake-up was not planned. The new task force will, however, report to Quayle and signals a continuation of his growing interest in controlling the space programme (see *Nature* 343, 310; 1990).

The past weeks have seen the discovery of flaws in the optics of the Hubble Space Telescope, the grounding of the shuttle fleet due to a mysterious fuel leak (see page 207), debate over a rash of apparent problems with the design of the Space Station, and congressional budgetary

EUROPEAN SPACE AGENCY

Remote-sensing satellite agreement

London & Paris

AFTER almost a year of uncertainty, the go-ahead has been given to build a second European remote-sensing satellite (ERS-2). At a European Space Agency (ESA) council meeting at the end of last month, France finally decided to contribute 23 per cent of the satellite's £297 million costs, bringing the total commitment from member states so far to 80 per cent.

At ESA headquarters in Paris last week, no one was willing to give a precise breakdown of funding for the satellite, as some partners, notably Italy, have not yet specified exactly how much they are prepared to commit. It is, however, expected that they will foot about 13 per cent of the bill.

News of the go-ahead for ERS-2 was greeted with relief within ESA. Last autumn there were fears that rising costs of the Columbus programme, including the Hermes shuttle and Ariane 5 launcher, would jeopardize the satellite. And France has its own Spot remote-sensing satellite programme, although with very different capabilities and aims. But ERS-2 is seen as vital to provide continuity with ERS-1, set for launch next spring, and the polar observation platform planned for 1997, part of the Columbus programme. ERS-2 is expected to be about 50 per cent cheaper to build than ERS-1 and will share some basic components with the French Spot satellites.

Peter Coles & Peter Aldous

assaults on several of NASA's projects.

"NASA is down but not out. We'll bounce back", says William Lenoir, associate administrator for space flight at NASA. A roller coaster ride is nothing new for NASA; following the low of the Columbia Space Shuttle crash in 1986, NASA came back with last year's triumphant Voyager flyby of Neptune.

NASA's future problems will be compounded by this week's news that the budget deficit for fiscal year 1991 is likely to be \$168,000 million, rather than the \$100,000 million predicted by President George Bush in January. To avoid triggering across-the-board cuts under the Gramm-Rudman Act, the White House and Congress must find ways to reduce the deficit to \$64,000 million — a task that will require huge cuts.

Action in the House of Representatives has already cut the president's budget request for NASA by \$800 million to \$14,300 million. Out went funds for a programme to search for intelligent extraterrestrial life — "let space aliens use their currency to find us", was the comment of Silvio Conte (Republican, Massachusetts) during debate in the House. Out also went funding for Bush's plan to land humans on Mars sometime early next century. Bush protested vigorously when cuts were suggested at the committee stage, accusing Congress of abandoning the American "pioneering spirit".

The House has also taken a tough stand on plans for the Space Station Freedom, taking the unusual step of writing a proviso into its appropriations bill that would cut off funding if power supply problems are not overcome by the year-end preliminary design review. The station is limited in the total power available to the 75kW supplied by its array of solar panels. But so much of that power is needed for house-keeping functions that NASA can guarantee only 15kW — a third of that originally planned — for scientists using the station.

Last week, Richard Kohrs, director of Space Station Freedom, said that the problem was not as bad as it seemed. Estimates of power demand had included excessive use of equipment such as microwave ovens. But other big savings still need to be found. Kohrs says that NASA is now examining ways of reducing the number of measurements of such things as temperature, pressure and air flow to reduce power demand.

Two other big problems with the space station worry Congress. By far the most intractable is that the station is now too heavy. The scale of the weight problems was revealed last week in a NASA report leaked to *Space Station News*, a Washington-based newsletter. The space station is 150,000 pounds over its total weight target

of 500,000 pounds.

Simply adding shuttle flights to build a heavier station would not win NASA points for strict management. Last week, Kohrs said that weight problems would be overcome and were not unusual at this stage of the design process. He presented graphs showing that the design weight of the station was now falling. If cuts can be continued at the same pace, the station will slim to the right weight just in time for the preliminary design review.

A new problem comes from NASA's most recent calculations of the time that astronauts will have to spend outside the station carrying out routine repairs. At first, only a couple of hundred hours of 'space walks' was thought likely to be necessary each year. A NASA report in February this year pushed the number up to over 2,000 hours. Now the figure is rising again. Space walks are considered dangerous and the effect of long exposure to space radiation is unknown.

Trouble is also looming for another of NASA's giant projects in a National Academy of Science's report, due out in mid-August, that will continue the debate over whether NASA should launch giant satellites as part of its Mission to Planet Earth (see *Nature* 344, 578; 12 April 1990). The mission, intended to understand global environmental changes, will include massive Earth Observing System (EOS) platforms, each carrying an elaborate suite of instruments. As confidence slips in NASA's ability to successfully manage large projects, the question is whether the platforms should be replaced by simpler satellites.

In principle, large satellites make more sense. They are cheaper to operate and can make the many simultaneous measurements, looking through the same air path, that are essential to model accurately the global environment. But the same data could be provided at less risk of a catastrophic failure by using a train of small satellites, following one another closely. Small satellites are also more flexible as frequent launches provide opportunities to add the most advanced sensor technology or to change instruments as new questions emerge.

More important, as pointed out in a briefing published last week by the American Institute of Aeronautics and Astronautics, clouds "cannot be studied effectively" by the EOS platforms. Yet long-term changes in cloud cover remain one of the key unknowns in predicting global warming.

NASA says that difficulties in large programmes have often arisen as a result of work being divided among several centres. Work on the Hubble Space Telescope was divided between two centres, that on the space station among four. EOS will be run entirely from Goddard Space Flight Center.

Alun Anderson

Congress get optics lesson

Washington

It was technical difficulty, rather than economic stringency, that persuaded managers of the Hubble Space Telescope (HST) project not to attempt an integrated test of HST's optical system on the ground, according to Lennard Fisk, associate administrator for space sciences at NASA (the National Aeronautics and Space Administration) in testimony given to the House of Representatives Committee on Science, Space and Technology last week. But Fisk admitted that with hindsight there were undoubtedly simple and relatively cheap tests that could have been performed on the assembled telescope which would have caught the image-distorting aberration in HST's mirrors (see *Nature* 346, 3; 5 July 1990).

Several members of the committee began last week's hearing worried that the decision to rely solely on pre-assembly tests of HST's primary and secondary mirrors demonstrated a penny-wise, pound-foolish philosophy on NASA's part. Even if an "all-up" test of the finished telescope would have cost hundreds of millions of dollars, they suggested, it would have been a worthwhile ingredient in a \$1,650 million project. But during three hours of questioning, Fisk, accompanied by NASA administrator Richard Truly, made a strong case that the decision to test the mirrors separately but not together was, if ultimately a mistake, nevertheless rationally made.

To achieve HST's desired sensitivity in the ultraviolet part of the spectrum, its reflecting surfaces had to be kept clear even of monomolecular layers of "polluting" chemicals. A giant clean room would therefore have been needed in which to carry out telescope tests. But more fundamental was the problem that HST was designed to work in zero gravity; with the primary mirror mounted in a frame designed for weightless conditions, gravity at the Earth's surface is enough to cause a distortion several wavelengths of light in magnitude, more than the half-wavelength error that has been discovered and much more than the hoped-for one-fiftieth of a wavelength precision of the mirror's profile.

Only by mathematically subtracting the gravity-induced distortion, a procedure which would itself need to have been checked and verified, could the mirror's true figure have been deduced from the results of a laboratory test. NASA decided that a test of such difficulty, and with questionable feasibility, was not worth doing.

Having concluded that no high-precision test was possible, project managers at the time evidently decided also that there was no purpose in performing any tests of

lower precision — tests that would not have verified that the optics were up to specification, but which would have found out the flaw that is now so painfully known to exist. This link in the reasoning, Fisk suggested, was something that the investigation into the cause of HST's aberration will examine closely.

Lew Allen, director of the Jet Propulsion Laboratory and chairman of the HST Optical Systems Board of investigation, said that records impounded from Perkin-Elmer, which made the mirror, and the Marshall Space Flight Center, which oversaw the contracts for the optical system, are being examined. If this paper chase fails, there are some pieces of hardware left over from the mirror-making process which can be examined. These too have been impounded.

The most important of these exhibits are likely to be three "null correctors". These are small optical assemblies of spherical surfaces through which laser light is shone on to the primary mirror to create an interference pattern; as the mirror is polished, the aim is to reduce the interference pattern from a collection of light and dark spots to a uniform illumination, at which time the mirror is supposed to conform to the desired shape. For this process to work, the null corrector itself must have precisely the right properties.

Two null correctors from Perkin-Elmer and one from Eastman-Kodak, which submitted a losing bid to make HST's optics but manufactured a partly finished back-up primary, still exist, and will be examined to assess directly whether HST's primary came out wrong because the null corrector used to control its final shape was incorrectly made. Cross-checks using the different null correctors on the back-up primary may also be done.

There also exists a companion secondary mirror, which was made in parallel with the one now in HST, the two of them being swapped repeatedly during final polishing and measuring. Because they are thought to be identical, Allen said, a measurement of the spare secondary at Lawrence Livermore National Laboratory should be a definitive test of whether HST's aberration is in the primary or secondary mirror. Analysis of recent images from HST suggests that the primary is at fault.

Allen was confident that his board would find the answers to the questions posed, although some of the optical tests might take several months to finish, but members of the Science, Space and Technology Committee were concerned that by aiming only to learn how HST's aberration came about, the board's purview was too narrow.

Fisk, trying to reassure the committee

Problems seem to be at an end

Washington

NASA (National Aeronautics and Space Administration) officials said last week that they are "98 per cent sure" that the two hydrogen-fuel leaks plaguing the grounded shuttles, the Columbia and the Atlantis, have different causes, that repairs should be easy and that the fleet should be flying soon. Space Shuttle director Robert Crippen said there is "no generic problem" in the design of the fuel lines that connect the shuttle with its external fuel tank which would have "been more difficult to deal with".

The leak on the Atlantis was detected as it was being prepared to launch a secret military payload in late June. Taking their cue from problems seen earlier with Columbia, NASA technicians at the time performed a pre-launch "tanking test" and found a leak. In an attempt to localize the leak, NASA workers last week put plastic bags around suspect joints and monitored them for the build-up of traces of hydrogen fuel. The tests revealed that almost all the hydrogen was leaking from a teflon-covered flange seal where the fuel assembly is bolted to the external tank.

Ending the leak may require only that bolts be tightened in order to squeeze the seal more tightly, according to William Lenoir, NASA space flight associate administrator, or it may require the shuttle to be rolled back into its hangar and a new seal installed. Early suspicions that the leak was at the point where the pipe disconnects when the fuel tank separates from the shuttle proved to be incorrect.

The leak on Columbia has already been fixed, said Lenoir. NASA workers replaced its fuel connection system with that from the Endeavor, a fourth shuttle which is still under construction. If the cause of the leak is found and curable, it may be possible to rebuild Columbia's fuel system and install it on the Endeavor.

Launch dates for shuttle flights are still unknown but priority is being given to the Atlantis. Discovery, carrying the joint US-European solar probe Ulysses, is expected to go up during the narrow time window in October in which the probe can receive a 'gravity assist' from Jupiter to speed it on its way to an orbit around the poles of the Sun.

Robin Eisner

that NASA was anxious to learn as many general lessons from the affair as could be found, said that Allen's investigation was simply the essential first step in what could turn out to be a much larger assessment of NASA's project management style. But the committee, especially chairman Robert Roe (Democrat, New Jersey), seemed to want more than a verbal assurance that Allen's findings would not be filed and forgotten.

David Lindley

Tissues not for sale

Washington

A CONTROVERSIAL 1988 California court decision that granted patients ownership of samples of tissues and bodily fluids taken from them was overturned last week in the state Supreme Court.

Had the lower court's decision been upheld, a patient (or a patient's heir) could have sued a researcher, institution or company for profits from any products created from the use of the patient's tissues; among them would have been many of the human cell lines in use in laboratories across the United States. In making its decision, the California Supreme Court abided by the tradition in Anglo-American law that maintains that an individual cannot sell his or her body parts. A person can, however, be reimbursed for the "service" of providing replaceable tissue, such as blood and semen.

The case began in 1984 when John Moore, who had cells taken from his spleen during treatment for hairy cell DNA FINGERPRINTING

No escape with snake DNA

New Delhi

INDIA'S first judgement in a paternity case made on the basis of DNA fingerprinting evidence has set a major precedent. Since the case was disposed of in April, even small courts in remote corners of the country have begun to rely on the technique in paternity disputes.

In India, paternity disputes make up a significant percentage of court cases. Many of the cases are now being referred to the Center for Cellular and Molecular Biology (CCMB) in Hyderabad where scientists have developed their own DNA probes. A test costs only Rs 3,600, about one-tenth of the cost in Europe or the United States.

The historic judgement was given by the chief judicial magistrate of Tellichery Town in Kerala State in a dispute between an unwed mother and her lover who denied having fathered her child. The judge ordered the man to pay for the maintenance of the child after hearing evidence from Lalji Singh of CCMB.

The judge dismissed attempts by the defendant to argue, on the basis of cases in the United States, that DNA fingerprint tests could go wrong. There is no specific Indian legislation on the admissibility of DNA fingerprinting evidence but 'expert' opinion is valid in court under existing law.

The CCMB probe, which is being patented, consists of a repetitive GATA sequence isolated from the DNA of the banded krait, a poisonous Indian snake.

K. S. Jayaraman

leukaemia, sued his doctor, a technician in the doctor's laboratory, the University of California at Los Angeles (UCLA), Sandoz Pharmaceuticals and the biotechnology company Genetics Institute, for profits derived from a cell line developed from the cells. The cells produce granulocyte-macrophage colony stimulating factor and other cytokines. The doctor and his technician had obtained a patent on the cell line and transferred it to UCLA, which in turn had granted licences to the two companies. For the transaction, the doctor received 75,000 shares now estimated to be worth \$3 million in Genetics Institute and UCLA received \$440,000 in research grants from the two companies. The companies have so far made no profits from selling the cell line.

The court decision will not end legal action in the case. Although the notion that patients can own their tissues was rejected, the court set legal precedent by ruling that a doctor, when seeking consent for a medical procedure, must "disclose personal interests unrelated to the patient's health, whether research or economic,

that may affect his medical judgment". If a doctor fails to obtain proper consent, the doctor and others party to this breach of conduct may be sued for damages. Moore's attorney is now suing the defendants for not properly informing the patient of possible commercial use of his cells.

Despite the continuing action, the defendants view the court's decision as a victory because it would have been easier for Moore to collect damages on a property claim than on a claim based on lack of informed consent. "We would only have had to prove taking of property", said Moore's attorney, Sanford Gage. Now he must show wrongful intent on the part of the doctor, when he failed to tell the patient of his interests, and show complicity on the part of the codefendants. The case is not expected to go to trial for another two years.

Gage may try to appeal against the property ruling in the US Supreme Court. But from now on, doctors in the United States will have to be more careful to inform their patients of their research or economic interests and ensure that their consent forms state that a patient has no claims on what happens to discarded tissues.

Robin Elsner

NEW ENGLAND JOURNAL OF MEDICINE

Editorial position vacant

Washington

ARNOLD Relman, editor-in-chief of *The New England Journal of Medicine* (NEJM), last week announced that he will retire next June.

Relman has been a controversial figure, often criticized by medical and hospital associations for his opinion that the profit motive in medicine has altered professional judgement and hurt patient care. The journal's strict policy of refusing to publish material that has been released to the press has also engendered harsh comment from the news media.

Reflecting upon his 14-year tenure, during which the journal's circulation rose by one third to 225,000, Relman said his greatest accomplishments were demanding greater originality in the manuscripts published by the 178-year-old journal, and making the journal a forum for discussion of the social, economic, political and ethical implications of medicine and medical research.

He thinks he also helped to "sharpen" standards of ethical behaviour in medical publishing. The journal was the first to demand that authors reveal "any commercial associations that might pose a conflict of interest with the submitted article". NEJM asks that funding sources, stock ownership and patent licensing arrangements be disclosed. Although a commercial interest does not disqualify a research manuscript from publication, Relman

instituted a new policy last July that prohibits an author of a review article or opinion piece from having a significant commercial connection with the ideas he or she is writing about.

To criticism of his strict rule of refusing publication of material that has been released to the press, Relman says that doctors cannot make decisions about patient care based on newspaper reports. "Doctors need to see the original data in the article", he said. Critics contend that the policy delays the release of information vital to the public. But Relman says that the journal makes an exception when a finding has been described in the press on the advice of a respected agency. Such cases include the release of information concerning the link between tampon use and toxic shock syndrome (on the advice of the Public Health Service), and of the value of prophylactic chemotherapy for women who had received surgery for localized breast cancer (on the advice of the National Cancer Institute).

On retiring, Relman says he will have time to develop his ideas and write books. He says he wants to tackle the issue of cost control in the "inflated" US health-care budget and what to do about the uninsured. Another subject is the dissemination of medical information in the media and its impact on the general public, with particular emphasis on the role of the peer-reviewed journal.

Robin Elsner

Reprieve for academy

Munich

THE East German Academy of Sciences is to be given an extra year to complete the massive process of restructuring required for it to be integrated into the West German scientific system. The academy, by far the largest research organization in East Germany, will now be given until the end of 1991 to finish its reforms, by which time a scientific review carried out by the West German science advisory council, Wissenschaftsrat, will be complete.

The West German Research Ministry (BMFT) agreed to the reprieve after the *Länder* and West German science organizations rejected Research Minister Heinz Riesenhuber's proposal to hack the academy down to size by the end of this year on the basis of only a rough evaluation by the Wissenschaftsrat (see *Nature* 346, 8; 5 July 1990).

In a strongly worded letter, Science Minister Marianne Tidick of Schleswig-Holstein argued that a quick change would be unfair to researchers and to the yet-to-be-created *Länder* in East Germany. By granting BMFT nearly dictatorial powers over East German research, the *Länder* felt an uncomfortable precedent would be set for West German

COMPUTERS

Hacker's genius award

Boston

RICHARD Stallman, a programmer who, at the age of 37, is well known as a colourful and strident advocate of free, public access to computer software, is among this year's recipients of the John D. and Catherine T. MacArthur foundation's 'genius' fellowships which are awarded annually to a disparate group of artists, writers, scientists and political activists. This year's awards have yet to be formally announced but Stallman, whose selection is sure to annoy the corporate heads of many large commercial software companies, acknowledged publicly his receipt of the \$240,000, five-year fellowship early this week.

In August, Stallman says he plans to lead a public demonstration against one such company, Lotus, Inc., based in Cambridge, Massachusetts to protest at a new lawsuit the company has brought against a competitor. The MacArthur award, he says, will help him step up his battle against what he calls "the new monopoly" on computer software, and allow him to support others working on "free software". Stallman, whose waist-length hair and beard fit closely his unorthodox, 'hacker' image, severed all formal contacts with Massachusetts Institute of Technology (MIT) in 1984 to ensure there could never be a claim of MIT ownership of any part of his work.

Seth Shulman

science policy.

In advance of its full evaluation, Wissenschaftsrat estimated last week that refurbishing East German universities and research institutes would cost a minimum of DM6,500 million (\$4,000 million) over five years. This amount is more than the entire budget of West Germany's Max Planck Society.

The East German Research Ministry estimates that it can cover just 20 per cent of the research costs but all the personnel costs of the academy for the remainder of 1990. The ministry implies that it may find more funds by using money originally earmarked for investment. How the new East German *Länder*, especially Berlin, will finance their universities is still a mystery. If Berlin is recreated as a single *Land*, as is currently planned, it will have three major universities.

Reports that the East Germany Academy would fire 14,000 of its 24,000 employees before the end of the year have turned out to be deliberate falsehoods — reminiscent of Communist disinformation of years past — released to the public in order to provoke a response from academy employees. The alleged cuts were leaked to the press and then dictated to institute directors at a meeting in East Berlin on 25 June. The strategy succeeded. Thousands of demonstrators gathered outside the meeting at the Adlershof research campus to protest against the threat to their jobs.

But just a day later, the Research Ministry denied that mass dismissals were planned. Only 3,000 to 5,000 people will be removed from the academy through early retirement or when institutes are turned into private companies.

Benno Parthier, director of an academy institute in Halle and the new president of the Akademie Leopoldina, said he expected the real number of layoffs to fall between the two estimates.

Wilhelm Krull of Wissenschaftsrat said he expects to receive a variety of recommendations on where to resettle academy researchers and institutes. There are many more options, he said, than just dissolving institutes, recreating them as Max Planck or Fraunhofer institutes (for basic and applied research respectively) or resettling them at universities. Some institutes might, for example, be turned into dedicated institutes in specific fields by the new *Länder*.

Bringing research back to universities in East Germany, from which it was forcibly removed 40 years ago, will be a long and arduous task, said Krull. At first, only isolated research groups will probably be moved, not least because research costs money and the new *Länder* will not be able to pay for it.

Steven Dickman

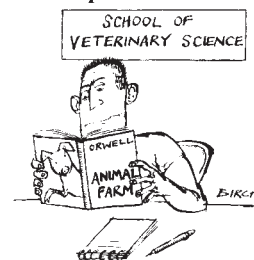
A plan for spending money

Sydney

IN a new report that may serve as a base for a reworking of the way universities are financed, the Australian federal government divides academic disciplines into five classes according to their teaching cost per student. Universities that are overfunded, according to the definitions, will be expected to take more students for the same total amount of money, and those that are underfunded will be allowed to shed students or will receive more government money.

The report, prepared by a joint committee of the Commonwealth Department of Employment, Education and Training and the Higher Education Council, puts accounting and the law among the cheapest undergraduate courses to teach and, as expected, finds agriculture, dentistry, medicine and veterinary science the most expensive, costing 2.95 times as much per student as the courses in the first category. Science and engineering occupy the fourth group, with 2.20 times the cost of the first.

The Federation of Australian University Staff Associations has strongly criticized the report. According to the general secretary, Di Zetlin, "calculations for this report have been based on only three teaching cost studies, all with small sample sizes and all coming up with different results". Zetlin claims that the report discriminates against "low cost/high infrastructure" disciplines such as those in the



humanities, and that faculties that cover areas in the high cost category will "do very well indeed at the expense of law, accounting and mathematics".

A separate report, issued recently by a Senate standing committee, criticizes the practice of employing university teachers on the basis of their published work rather than their educational abilities, with the result that Australian universities and higher-education colleges are turning out "skilled barbarians." The report, "Priorities for Reform in Higher Education," found that 40 per cent of graduates proceeding to a diploma in education failed simple literacy tests; science and medical graduates were "culturally illiterate," completing their courses without having written a substantial piece of English; and fewer than one in three humanities students could summarise a passage of prose by George Orwell.

Tania Ewing

Foreigners join MITI's club

Tokyo

LAST week, a large group of senior officials from Japan's Ministry of International Trade and Industry (MITI) held a first informal meeting with a handful of foreign science correspondents based in Tokyo in an attempt to open new lines of communication between the ministry and the foreign press.

The meeting may not seem like a historic event. But in Japan, it marked an unprecedented breakthrough that should allow more open coverage of MITI's activities in science and technology. Bob Johnstone, science and technology correspondent for the *Far Eastern Economic Review*, and several MITI officials organized the meeting. Like many foreign correspondents in Japan, Johnstone is frustrated by the lack of information provided by government organizations.

At the meeting, 25 senior officials from MITI's Agency of Industrial Science and Technology (AIST) promised to open new channels of communication by providing press releases, holding press conferences for the foreign press and, more important, by arranging regular, informal meetings with senior AIST officials.

At present, almost all government information is fed to exclusive private press clubs (*'kisha'* clubs) connected to each ministry or organization. Until recently, nearly all of these clubs were closed to foreigners. Although the Foreign Press Club in Tokyo has led a long campaign to open all such clubs to foreign members, the advantages of joining may be questionable.

Japanese *kisha* club members virtually live in the organizations they cover. The clubs are located on the ministry premises and beds are provided for journalists who want to stay overnight.

Because reporters from many (but not all) newspapers are assigned to particular clubs for long periods, there is a strong incentive for them not to upset the organization in which they are located.

As a result, the clubs sometimes seem to be more like public relations departments than press organizations and critical review of government policy may be lacking.

MITI officials will be disappointed if they expect the new 'club' of foreign science correspondents to behave like *kisha* club members. But MITI seems well aware of the differences between the foreign and Japanese press.

Masaru Sugiura, director-general of AIST, in his opening address to foreign correspondents, noted that "the value of information is often construed as different between the party retaining and distributing such information and the receiving party thereof" and "that this is particularly

true between the Agency and the foreign press". Nevertheless, he asserted that "information gaps that may exist between us should be immediately dissolved, and information on Japan's science and technology policies should be given properly to foreign press representatives and correspondents".

MITI's move to open new communication channels comes at a time when the ministry is actively trying to international-

PUBLIC HEALTH

Bad blood between Red Cross and FDA

Washington

THE American Red Cross came under fire last week from the US Food and Drug Administration (FDA), which issued a preliminary report claiming that several hundred instances of errors or accidents in the collection and distribution of blood were not reported promptly to the FDA but languished instead at national Red Cross headquarters in Washington, DC.

At a hearing on the safety of the nation's blood supply, a congressional subcommittee on oversight and investigations, chaired by Representative John D. Dingell (Democrat, Michigan), set out to determine whether the lessons of the 1980s, when tardiness in dealing with HIV (human immunodeficiency virus) contamination shook public confidence in the blood supply, had been heeded by the blood banking industry and government regulators. As Dingell said, the inquiry "focuses on whether the blood supply is as safe as it could be or should be".

The subcommittee heard FDA charges that the American Red Cross, which supplies about half of the nation's blood supply, either failed or was slow to report to the FDA 386 error and accident reports. The FDA requires that all such reports be passed on promptly, which means, according to Mary T. Carden, one of the investigators responsible for the FDA's report, in "10-15 days". Carden noted one case where plasma with elevated levels of alanine aminotransferase, an indicator that the blood may harbour other viral infections, was erroneously shipped to the Swiss Red Cross and not reported to the FDA for two years. One Red Cross employee, when questioned about delays, had said "error reports were not a first priority".

Lewellys F. Barker, senior vice president of the American Red Cross, said that the 386 reports "are in the process of being reviewed", and that "when a region discovers an error, they correct it immediately". He said that the Red Cross will provide a written response to the FDA's allegations in 60 days. Diane Gershon

ize its major research projects by encouraging foreign companies and researchers to join them.

It is to be hoped that other ministries and government organizations will follow suit. Japanese scientists often complain that their research does not get the recognition it deserves in the West. But those scientists who are in universities and government research laboratories need only look as far as their *kisha* clubs to understand why they fail to get wider coverage in the Western press.

David Swinbanks

KANSAI SCIENCE CITY

High-energy ions for rent in Japan

Tokyo

ONE of the first of several new research institutes planned for the new Kansai Science City being built not far from the cities of Osaka, Kyoto and Nara was opened on 5 July. The Ion Engineering Center is equipped with the world's most powerful ion implantation apparatus (8 MeV), and will 'rent' its facilities to both private and public users. The centre provides one example of how organizers of the science city are attracting private investment.

The centre is managed by a semi-public corporation established in 1988 by the New Energy and Industrial Technology Development Organization (NEDO) with an initial investment of ¥7,800 million (\$52 million) from NEDO, the Osaka, Kyoto and Nara prefectural governments, and several private companies, including Kansai Electric Power Company, Nippon Steel, Hitachi, and Matsushita Electric Industrial Company.

Private companies, universities and national research institutes and laboratories will be able to rent the centre's ion implanter and ion deposition apparatus for research on new devices and materials, including solar cells, fuel cells, high-performance sensors, superconductors and biomaterials. A separate company, the Ion Engineering Research Institute Corporation, will also conduct its own research at the centre.

The centre joins the Advanced Telecommunications Research Institute International (ATR), the first institute to open in the science city (see *Nature* 338, 285; 1989). ATR is backed by the Ministry of Post and Telecommunications and is financed with funds derived from the dividends of government-held shares of the semi-privatized domestic telecommunications company Nippon Telegraph and Telephone (NTT). When completed early next century, the science city is expected to attract a total investment of \$25,000 million dollars, much of it from the private sector.

David Swinbanks

Paying the price of peace

Boston

US university programmes for the study of arms control and international security may be becoming casualties of the very decrease in superpower tensions that many of the programmes' researchers have long advocated. While some of the directors of academic programmes in arms control and security studies around the United States say that they have yet to experience a downturn in funding, they are virtually unanimous in their expectation of such a drop. Many have already begun to respond by seeking to broaden their missions. As one researcher in the field put it "everything on my bookshelf is out of date."

Harvey Sapolsky, who heads the Defense and Arms Control Studies Program at the Massachusetts Institute of Technology (MIT), echoes the sentiments of many of his counterparts when he says that the private foundations that have traditionally funded the field tend increasingly to "see security issues as being solved" by recent world developments.

BSE

No madness in UK policy

London

UK agriculture minister John Gummer's often-repeated assertion that "beef is safe" is endorsed in a report from an all-party House of Commons committee investigating the epidemic of bovine spongiform encephalopathy (BSE) in British cattle, published last week. The agriculture select committee also concludes that the UK government's reaction to BSE represents a "substantial improvement" on its handling of the 1988 salmonella outbreak in British eggs — that food scare forced the resignation of a junior health minister and brought criticism from the agriculture committee.

But the report suggests that some action, notably banning certain cattle offals for human consumption and providing full compensation to farmers for BSE-afflicted cattle, could have been taken more quickly.

In a departure from tradition, the committee singled out one witness for attack. Professor Richard Lacey, a University of Leeds microbiologist, told the committee that "we could virtually lose a generation of people" through BSE. The report describes his evidence as "a mixture of science and science fiction", and criticizes the UK media for the extensive coverage given to "the pronouncements of one or two irresponsible scientists", which fuelled public anxiety over the risks to human health.

Gummer has rightly relied on advice

"On the contrary", Sapolsky argues, "we think that many of these issues are more exciting than ever now as the theology is challenged and forty years of accepted wisdom is overturned."

Sapolsky says he hopes foundations realize the extent of "important work" left to be done in the field — questions about "what in the military should be shrunk and by how much", and assessments of potential "threats posed by the new international circumstances". Sapolsky stresses that his programme, like others, has begun to consider a broadened notion of security. Current plans include a proposed collaboration between environmental scientists, lawyers and security experts.

The reduction in funding for university programmes in arms control and security issues will mark the end of more than two decades of growth. In the United States, private funding for these programmes began in the late 1960s when the Ford Foundation, under the direction of former White House security adviser McGeorge Bundy, provided several large grants.

from an independent committee of scientists "rather than negotiating with vested interests", the report says; but he should be prepared to go beyond those measures justified on scientific grounds alone, to reassure the public or for commercial reasons.

The report goes on to recommend a number of further precautionary countermeasures. The ban on human consumption of specified offals should apply to all cattle and splitting of cattle heads to remove the brain, should be banned. Although practised in only a small minority of UK abattoirs, several witnesses to the committee were concerned that this could spray an aerosol of potentially infective brain tissue over the head meat.

Other suggested countermeasures are based around the possibility that BSE may be transmissible vertically, from cow to calf. Calls for a ban on breeding from the offspring of BSE cases are rejected, consistent with advice from the government's independent scientific advisory committee. But the report suggests that this breeding should be discouraged: if BSE occurs as a result of vertical transmission, farmers should not be eligible for compensation. To ensure that the offspring of BSE cases can be tracked down if vertical transmission is shown to be possible, the report recommends that a central computer record of cattle breeding and movement should be set up (see *Nature* 345, 277; 24 May 1990).

Peter Aldhous

Many of the same foundations are shifting their focus away from traditional studies of defence and arms control towards the environment and development issues in Eastern Europe and developing countries.

But is current confidence in international stability well grounded? John Mearsheimer, who heads the political science department at the University of Chicago, predicts that the coming fall in funding will be only a "brief lull" based upon "false optimism" about global stability. Issues such as the potential proliferation of nuclear weapons in Europe, Mearsheimer predicts, will create "very serious trouble" for international relations and are likely to renew foundation support to the field. In the meantime, Mearsheimer calls the inclination of many of his colleagues to broaden the definition of security to include ecological and other concerns "an assault" on the field.

Paul Walker, director of the Institute for Peace and International Security based in Cambridge, Massachusetts, says that the funding situation is presenting immediate problems for smaller organizations. Walker expresses dismay that foundations are turning to other areas because the current situation presents a "marvelous opportunity" to establish new security arrangements. Major cuts in defence spending and uncertainty over the future of the North Atlantic Treaty Organization will require a fundamental rethinking of military roles, he says.

Walker, who is also a graduate of MIT's arms control programme, complains that the current funding climate has also fostered a willingness to approach "inappropriate" funding sources. MIT's programme, for example, has appealed to defence contractors for general support for the first time in its history. Sapolsky claims that such funding will not hamper the programme's independence, but Walker calls the move "outrageous," likening it to "lung cancer researchers appealing to cigarette manufacturers" to support their work.

Coit Blacker, a professor of international relations at the University of Southern California and former associate director of Stanford's Center on International Security and Arms Control, acknowledges many of the problems, but puts them in a more positive light. Blacker stresses that those programmes that are "most responsive" to current shifts in the priorities of the field will be best able to survive the current funding climate. Like Mearsheimer, Blacker espouses the view that the past four decades have presented "an atypically quiet period in world politics" because of the relatively stable, "bipolar competition" between the United States and the Soviet Union. Now, with an increasingly "multi-polar world", he says, "the potential for international violence is increased".

Seth Shulman

Stewardship of resources

SIR—Your leading article "Doctrinal fallacies of stewardship" (*Nature* 344, 179; 1990) claims that the loss of various natural resources through consumption by commerce and industry have been "more than compensated for by gains in social capital (which term includes not only most countries' educational systems but the world's accumulated stock of technical skill)". This is highly debatable. And it is certainly not immediately obvious that even all of the article's seemingly random list of lost resources (auks, emeralds, surface coal, molybdenum (soon) and amber), have contributed to social capital in due proportion to their original quantity or value. Have auks or amber done so? It is in this light that the scorned remarks of M. Jacques Delors, president of the European Commission, concerning the need for respect for "nature for itself", should be viewed as an appropriate and balanced warning that things — including living creatures — are all aspects of one reality that civilized persons ought to respect.

As for the "horrendous pollution of Eastern Europe" being regarded as a "product of economic stagnation", it should be noted that industrial pollution, so prevalent and widespread in the West, is especially severe in countries that are customarily regarded as examples of progressive and dynamic economies. The real point is surely that criminal industrial mismanagements have been equally responsible for pollution in both East and West. And are we to suppose that the article is serious when it claims there is no place "where the impairment of public health by industrial pollution outdoes the improvement of public health brought about . . . by the rational improvement of water supplies and . . . by the practice of . . . often rudimentary medicine"? What an absurd notion of a rational trade-off this would be. Anyway, even if the measures begun a century ago did clean up domestic water supplies in the West, the acceptability of these same supplies is now everywhere threatened by present industrial pollution from which further "rational improvement" appears unlikely to give general (and economical) further relief. Future corrections are likely to require far more radical approaches.

As for petroleum, it makes no sense, of course, to conserve it, if the future world will not need it. But what if our descendants could find different and better uses for petroleum, that we are unable to foresee? Meanwhile, environmental scientists are distressed not by loss of petroleum but by the effects of automobile exhausts on quality and temperature of the atmosphere.

As for species exterminations, how can

the existence of species that man has "made to flourish" offset the horrendous loss of species already logged — losses that continue? What of the consequences of reduction of genetic diversity and in the variety of organisms, as these will affect the life sciences? Is the imminent demise of the African elephant, the rhinoceros and so on to be viewed with calm?

Your writer approves the United States' proposed elimination of its budget deficit "that cramps its freedom" by "growing so fast that the tax-yield will grow faster than the government's obligations". This must be a new "trickle down" economics that may be presumed will eventually benefit developing countries as US industry pays them the usual Western pittance for their labour and for further denudation of their unrenewable resources.

Rate of change — of the environment, of industrial activity, of human population growth — is the rock we are all apt to perish on. And that problem, caused by heedless governments and serviced by an industrial machine that runs without a governor, is what we find increasingly difficult to respond to. We can — and must — still win.

A. H. WEATHERLEY

*Division of Life Sciences,
University of Toronto,
Scarborough, Ontario
Canada M1C 1A4*

Abortion limits

SIR—It is in fact not true that the "legal time-limit for abortion" in Britain is at present 28 weeks ("Abortion from a hat", *Nature* 344, 476; 1990), even though this is now widely believed.

What the Infant Life (Preservation) Act 1929, explicitly confirmed by the Abortion Act 1967, actually says is, first, that "any person who with intent to destroy the life of a child capable of being born alive, by any wilful act, causes the child to die before it has an existence independent of its mother" shall be guilty of the offence of "child destruction". And, second, that "evidence that a woman had at any time been pregnant for a period of 28 weeks or more shall be prima facie proof that she was at that time pregnant of a child capable of being born alive". That, incidentally, must mean "capable of surviving if born alive".

But this neither says nor implies that before the 28th week a child must be presumed not to be capable of being born alive, and nowadays a healthy prematurely born baby, if given all proper care, has an increasingly good chance of surviving from about the 22nd week onwards. It would seem therefore that to destroy any unborn child at this stage would be an

offence under the 1929 Act, if strictly interpreted.

C. B. GOODHART

*Gonville & Caius College,
Cambridge, UK*

■ Under the Human Fertilization and Embryology Bill, Members of Parliament have voted for an upper limit of 24 weeks, except where the mother's life is endangered or the fetus is handicapped. □

Embryo research

SIR—Peter Aldhous (*Nature* 345, 283; 1990) alleges that "the present Interim Licensing Authority guidelines recommend that three IVF embryos should be placed into the uterus to ensure a reasonable success rate". The guidelines say no such thing.

In its fourth report (1989), the Interim Licensing Authority drew attention to the most recent statistics (from 1987) which showed the large incidence of multiple births resulting from IVF (*in vitro* fertilization) and GIFT (gamete intrafallopian transfer) treatment, with consequential medical, social and economic disadvantages. The relevant guideline (12) states that "consideration must be given to ensuring that, while a woman has the best chance of achieving a pregnancy, the risks of a large multiple pregnancy are minimised. For this reason whether IVF or GIFT procedures are used either jointly or separately no more than three eggs or pre-embryos should be transferred in any one cycle, unless there are exceptional clinical reasons when up to four may be replaced per cycle." This guideline sets *upper limits*; otherwise it makes no recommendation as to the number of pre-embryos to be transferred, but it warns of the potential dangers.

G. S. DAWES

(Deputy Chairman,

Interim Licensing Authority)

*6 Belbroughton Road,
Oxford OX2 6UZ, UK*

SIR—Your readers will wish to learn that among the countries with legislation on embryo research (*Nature* 344, 691; 1990) is Spain. In November 1988, the Spanish Parliament passed a bill titled *On Techniques of Assisted Reproduction* (BOE 282, 33373-33378, 14 November 1988), which makes it possible to carry out any procedures aimed at assessing viability of live preembryos, described as fertilization products up to 14 days, as well as at detecting a hereditary disease and advising against their transfer for procreation.

JOSE ANTONIO ABRISQUETA

VITALINO ALLER

*Centro de Investigaciones Biológicas
(CSIC),*

*Laboratorio de Genética Humana,
Serrano, 113 bis, 28006 Madrid, Spain*

Managerial mediocrity

SIR—Walter Bodmer's defence of the corporate plan for the Natural History Museum (*Nature* 345, 509; 1990), makes strange reading, and does little to allay fears that a disaster is looming. He chides some of us for not having "the time or the inclination to read the Corporate Plan carefully", but acknowledges himself that it is only an outline. More importantly, nowhere is any justification given for the closure of certain areas. Why (it is said) are studies of parasitic worms (human health), gems and building stones (mineral resources), archaeozoology (human origins), diatoms (environmental quality) and fossil plants (biodiversity) to be stopped and on the authority of what expertise?

Bodmer stresses that the museum is not immune to "stringent periodic review by distinguished assessors". How very curious that the corporate plan escaped such valuable external guidance. But the plan is much more fundamental to the future of the museum than just loss of posts. It heralds a radical reorganization. The museum appears to regard the plan as a *fait accompli*, and nowhere do I read of real flexibility, let alone dialogue.

The corporate plan is now contentedly grazing in the lush pastures of managerial mediocrity. Morale in the museum is at rock bottom. What Bodmer refers to as an "enormous response" is almost entirely critical of the plan. He writes also that "we have no alternatives but to live within our means". Nobody doubts this. But means are relative, and the promise of European support linked to the great museums in France and Germany needs urgent exploration. At the moment we had better write c/o Natural History Mausoleum, and leave the final comment with the chief trustee who invites us "to learn . . . from the wonders of Disneyland". Words fail me.

S. CONWAY MORRIS

Department of Earth Sciences,
University of Cambridge,
Downing Street,
Cambridge CB2 3EQ, UK

SIR—It is widely held that there is only one way to do science, this way being paradigmatically that of physics. There are consequently phenomena such as what has been called 'physics envy', where a science is distorted by some to conform as closely as possible to physics, and a *scala scientiae*, where sciences are ranked from physics at the top to systematics at the bottom.

Indeed, systematics is not physics. It is also not stamp collecting, with which it is sometimes more or less equated, initially by a physicist. Papers in systematics are written so that their surface is legible to

nonsystematists. Their real meaning is not thereby made intelligible.

The meaning of systematics is the underlying structure and form of evolution². This is not a trivial problem; its astronomical equivalent is the structure and history of the Universe. That the diverse methods of systematics do not approximate to those of physics should be a source of enlightenment rather than disdain.

For science really is done in very different ways. The basic methods predominant in systematics are historical and comparative; Gould³ has recently celebrated their significance. And there are reasons for these methods. It is ironic that systematics should be savaged at a time when its literally dependent offspring, comparative biology, palaeobiology and conservation biology, are ascendant.

It is also ironic that systematics celebrates diversity but is itself a victim of stereotyping and the narrow minds of others. For many years, the Natural History Museum in London was overall the leading centre of systematics in the world. This function of the museum is not for exhibition; it is not for service to others. The museum is, extraordinarily, denigrating the very science in which it has been pre-eminent in order to take on the flashy and the faddish.

LEIGH M. VAN VALEN

Department of Ecology and Evolution
and Committee on the Conceptual
Foundations of Science,
University of Chicago,
915 E. 57 Street,
Chicago, Illinois 60637, USA

1. Cohen, J. E. *Science* 172, 674–675 (1971).

2. Van Valen, L. M. *Nature* 323, 664 (1986).

3. Gould, S. J. *Wonderful Life*. (New York, Norton, 1989).

Censors leave gaps

SIR—In the survey on the state of science in Eastern Europe (*Nature* 344, 599; 1990), transfer of information is justly mentioned as one of the prerequisites indispensable for future science policy in the countries involved. The dearth of hard-currency literature is really crippling, and in view of our economic problems, as well as ever increasing prices of Western books and journals, the prospects for the near future are not bright.

We do not share, however, the pessimistic view that decades will pass before information technology in these countries makes affordable general access to journals abroad. One promising project in this field is already under way: the SatelLife library project for linking national centres of Eastern Europe with the British Medical Association Library.

One specific problem we were faced

with in the 1970s and 1980s should, we hope, remain only a remembrance of the bitter past: censorship of Western scientific journals. Hundreds of expurgated issues are missing, and gaps in collections of Czechoslovak academic libraries have to be filled up. The favourite victims of censorship on the list of periodicals to which the National Medical Library in Prague subscribed were *Nature*, *Science*, *British Medical Journal* and *Lancet*.

JIRÍ DRBÁLEK

Ústav vědeckých lékařských informací,
Sokolská 31,
121 32 Prague 2, Czechoslovakia

Second class scientists?

SIR—In a recent issue of *Nature* (26 April 1990, Classified 11), the University of Auckland advertised a Senior Lectureship in Medical Molecular Biology. The salary range was 30–43 per cent higher for medically qualified incumbents than for those not medically qualified: What possible purpose can such a discriminatory difference in salaries serve?

No doubt the University of Auckland, along with many similar institutions, feels that it needs to offer high salaries to attract candidates with medical qualifications away from more remunerative activities. If so, this should set the rate for that position. On review of the applicants, the appointments committee may find that a candidate without medical qualification is, nevertheless, the best qualified for the position. Subsequently to offer that candidate a substantially lower salary because he or she failed to come through the medical profession is a clear insult. At the very least, it sends a sharp message about the level of respect a non-medically qualified incumbent can expect from his medical colleagues.

The pity is that most self-respecting scientists without a medical qualification are unlikely willingly to face such a gratuitous insult. This can only be a loss to any medical faculty that truly seeks the broad spectrum of ideas and interest essential to the pursuit of science and teaching.

In such circumstances, one has to wonder what the term 'equal employment opportunity employer' means. In the present case, it presumably indicates an undertaking of fair treatment in being offered employment, but no guarantee of fair treatment thereafter. Scientists without medical training are very clearly to be considered second-class citizens.

JOHN P. GIBSON

Ontario Agricultural College,
Department of Animal and
Poultry Science,
University of Guelph,
Guelph, Ontario, Canada N1G 2W1

A view on tropical deforestation

Ricardo Radulovich

Tropical deforestation is frequently decried as a potential cause of future environmental disaster. But is this attitude fair? And does it take into account the reality of life in the developing countries?

EVERYONE is aware of the campaigns throughout the world to preserve tropical forests, mainly rainforests. The arguments in favour of preservation, as well as simple contempt for wastage and the loss of natural wonders, fall into two main groups: one relates to global concerns, including the greenhouse effect, biodiversity and worldwide availability of timber; the other addresses more local concerns, such as soil erosion, hydrological cycles, low agricultural productivity in deforested lands and shortages of timber and wood for fuel.

Given that the problems faced by tropical countries are many more and are often more severe and urgent than those associated with deforestation, it must be the set of arguments related to global concerns which triggers and supports campaigns against tropical deforestation. Of course, as local actions affect deforestation in the countries where they occur, these actions and their local effects also come under scrutiny from outside.

When dealing with problems other than those exclusively related to the developing countries, several questions are raised. How important can the greenhouse effect, biodiversity or future worldwide timber shortages seem to a third-worlder, for example? It would be wrong to think of this third-worlder as irresponsible or as a squanderer of resources before serious steps are taken towards solving the well-known and much worse problem of resource squandering and environmental irresponsibility in the developed nations.

Global effect

The contribution of deforestation to the yet unproved greenhouse effect¹ has been calculated to be at most a fraction of that of combustion of fossil fuels (petroleum, gas or coal) which occurs primarily in developed nations. This calculation does not take into account other potential contributors to the greenhouse effect, such as trace gases². Even if forests were to disappear before the end of the century, the effect on the environment will not accrue any further. Yet, during this time and for decades afterwards, combustion of fossil fuel will continue. Thus, in this context, deforestation will in retrospect have contributed only a minimal increase in atmospheric CO₂.

There are already programmes to preserve biodiversity in the tropics; only a small fraction of the land area of each

country or region is required to support the vast majority of species — after all, how much land for preserving biodiversity can a poor country afford and manage? One reason why there is pressure from outside to preserve areas larger than necessary could be because of the enormous economic potential offered to multinational industries by tropical plant species. Another issue is the future availability of tropical timber for world markets. But it is precisely the developed countries and their multinational companies who are promoting this depletion — aided, naturally, by their local partners.

So what motivates this tremendous campaign against deforestation? It would be naive to think of it simply as altruism awakened by the unnecessary loss of natural riches, when one considers the powerful political and economic forces that shape and rule the world. So, looking beyond the obvious, it may be asked, for example, whether there is an intent to deprive developing countries of the opportunity to develop large land areas into productive agricultural schemes. An excellent example of such a scheme is that of the Cerrado region in Brazil, where several million hectares have now been put into production. At present, soybean yields nearly equal those of the United States, and more than two crops can be obtained per year if irrigation is available³.

Regarding the sustainability of agriculture in the tropics, centuries of terraced rice production in Asia⁴ and consecutive sugar-cane cropping in the Dominican Republic and other regions⁵, show that even very steep land and acid-infertile soils can be productive indefinitely. Over-simplified statements, such as comparing the worth of a single hamburger to half a ton of rainforest⁶, as an argument to deter deforestation, are clearly invalid if farming with adequate technology can take place on deforested land. Moreover, new practices could be developed much faster if more reliance is placed on tropical agriculture. One example of this is related to coping with the all-too-common soil phosphorus deficiency. Important recent achievements include the elucidation of mechanisms for efficient plant uptake of soil phosphorus⁷ and the optimization of long-term use of fertilizer phosphorus⁸.

Why is there so much support for the many academics, conservationists and others involved in the preservation of rainforests? Could such preservation be of

direct benefit to other groups — perhaps to the large polluting industries of developed countries, or even to consumers — so that an exaggerated portion of the blame for worldwide environmental problems is placed on tropical countries? What proportion of this support is derived from chemical companies interested in using tropical rainforest species?

Perhaps, from the point of view of the human species, it is simply intolerable to lose the rainforest as a resource for the future, as its loss would be forever, whereas other problems in the tropics, like enormous daily losses in quantity and quality of human lives, can wait to be solved as human populations will always be replaced.

Serious problems

There can be no doubt in the mind of any reasonable person that the squandering of natural resources is not only foolish but even criminal, and it must stop. But so must other, more serious and urgent problems in tropical developing countries, such as poverty and malnutrition. These problems will not be solved by better management of forests, no matter how beneficial in economic terms such management may be. For local officials concerned with local problems, deforestation must come lower on the list when compared with other problems.

Because the severity of many problems in the tropics is so large when compared with deforestation alone, altruism is an unacceptable excuse for people outside these countries to demand an end to deforestation. Those who have interests at stake or serve the interests of others cannot expect their advice to be accepted, especially if the problem of deforestation is considered in isolation, ignoring the other realities of life in the third world. □

Ricardo Radulovich is in the Department of Agricultural Engineering, University of Costa Rica, San Jose, Costa Rica, and the Department of Soil, Crop and Atmospheric Sciences, Cornell University, Ithaca, New York 14853, USA.

1. Reifsnnyder, W.E. *Agric. For. Meteorol.* **47**, 349–371 (1989).
2. Lashof, D.A. & Ahuja, D.R. *Nature* **344**, 529–531 (1990).
3. Abelson, P.H. & Rowe, J.W. *Science* **235**, 1450–1451 (1987).
4. De Datta, S.K. *Principles and Practices of Rice Production* (Krieger, Malabar, 1987).
5. Blackburn, F. *Sugar-Cane* (Longman, London, 1984).
6. Uhl, C. & Parker, G. *Interciencia* **11**, 214 (1986).
7. Ae, N. *et al. Science* **248**, 477–480 (1990).
8. Smyth, T.J. & Cravo, M.S. *Agron. J.* **82**, 305–309 (1990).

What to do with extraneous data

This journal intends to follow some others in making available data that do not strictly belong to the scientific papers which it publishes. But that could change.

WHEN is a scientific research report, called a "paper" since the genteel eighteenth century when papers were "read" (out loud, before a small group of interested and disinterested parties), complete? When a fellow scholar, after a few careful readings of the text, is persuaded that he can so fully understand what has been done, and why, that he could march to his own laboratory and repeat the observations he has seen described with a fair chance of replicating their authors' results successfully. Naturally, this does not often happen. The more complete a paper, the less likely are its readers to strive at its replication. But the principle supervenes: to be credible, a paper must contain the essence of its replication. On that, everybody agrees.

The difficulty in that general statement is that the definition of completeness is like the definition of the length of the ideal piece of string: it depends on what the string is needed for. Most papers published in the journals of record (of which this is one) give well-informed readers a good sense of how they might most usefully set about the task of replication; there are densely printed figure legends that detail how the work has been done, and which are therefore a guide to the ways in which it might be repeated. Sadly, of course, the people best placed to understand these coded descriptions of experimental protocols are most of all interested not in replication but in going one step further.

At the same time, it is conventional that the author or authors of a paper will select from the whole corpus of the data at their disposal the few items that will help to prove their case. In palaeontology, for example, people will describe a single specimen in immense detail, and send a dozen others off to a museum unrecorded, on the indisputable principle that a single well-recorded specimen has greater value than sketchy descriptions of a greater body of material.

In human genetics, it is now commonplace to illustrate a general principle by a single example of a familial pedigree — and to put the others in a bottom drawer. In molecular biology, people seem to know that journals will no longer print nucleotide sequences *in extenso*, with the result that they offer for publication only the sub-sequences that, perhaps by comparison with other sub-sequences, will convince percipient readers that there is a case to be made for a particular molecular

mechanism.

Potentially, these practices are a means by which rafts of data may be permanently lost to the research community. Everybody knows that the best that can happen to data that have led nowhere is that they should be "archived", which means in practice that they should be recorded on magnetic tape and sent to a national or international centre for storage indefinitely.

But storage is expensive, and also hazardous; the General Accounting Office in the United States estimated the other day that 10 per cent of the data gathered by the US National Aeronautics and Space Administration (NASA) had been lost because the magnetic tapes concerned had been kept at too high a temperature, or simply misplaced. And that in a field in which the mechanics of data-gathering have been given close attention.

That is why this journal has embarked on a scheme for assisting with the publication of data that would otherwise be buried in people's desk-drawers. For the past few weeks, editors responsible for particular manuscripts submitted for publication have been offering to their authors the opportunity of submitting extra material for publication by a novel route; readers who get to hear, usually by means of an announcement in *Nature*, that there is something extra on offer will be invited to write in, and will be offered a simple photocopy of whatever the author or authors may have supplied, in the first instance without charge.

The immediate beneficiaries of this scheme should be the molecular biologists generating nucleotide sequence data, only some of which is relevant to the conclusions they wish to draw. Why should they not send in the extra data on the understanding that it will be published informally to those who wish to know what it consists of, alongside the more general circulation of their main message?

But there are other fields in which a service such as this should be helpful not merely to the world at large but to those who publish scientific papers — astrophysicists are forever finding that they cannot publish lists of all the stars they have observed, seismologists are at a loss to know what to do with compilations of seismic data, ocean drillers' logs of the constitutions of deep-sea cores hardly ever see the light of day. But why should they not be generally available?

That is the principle on which *Nature*, in

its dealings with those who kindly contribute its scientific articles, will in future function. Those contributing articles accepted for publication will be asked — and are hereby invited — to submit for informal publication such supplementary information as there may also be to hand. *Nature* will undertake to make that supplementary information permanently available. Often, there will be occasions when referees will suggest that the authors of disputable or even contentious articles should, in prudence, put supporting evidence through the same route.

For the time being, at least, none of this should suggest that anything will change. Authors who say that their supplementary information seems to them a perfectly satisfactory basis for a second bite at the same cherry in another journal will not be pilloried on that account. Nor, for the time being, will it be required of authors that they should replace long descriptions of experimental techniques by references to some standard handbook, on the understanding that a full account will be distributed to other parties sufficiently interested to ask for it.

For the time being, the mechanics of the process will be simple, and as described above. Contributors and would-be contributors need not now change their ways. But in the long run, they may discover a novel pressure breathing down their necks; not that expected of even more recalcitrant editors, but that stemming from the inner voice that asks whether this or that typescript will have a better chance if some of the data it reports are tucked into neat supplementary data sheets. If a large proportion of this journal's present contributors followed such a course, *Nature* would be able to publish a greater diversity of material.

The more important objective is that editors and contributors should come to a more direct appreciation of each other's often conflicting interests. Most editors have an interest that most readers should understand as much as possible, most contributors that what they have to say should be unassailable even at the cost that it is unintelligible.

That is why contributors (and readers, often the same people) are asked to give this venture an honest trial. Eventually, of course, supplementary information will be distributed electronically, through an electronic database. But that is light-years away.

John Maddox

What makes a man a man?

Anne McLaren

THE hunt for the male-determining factor in mammals has been both exciting and frustrating. Three reports on pages 240, 245 and 279 of this issue¹⁻³ mark what could be the end of the search, and the beginning of the genetic analysis of mammalian sex determination.

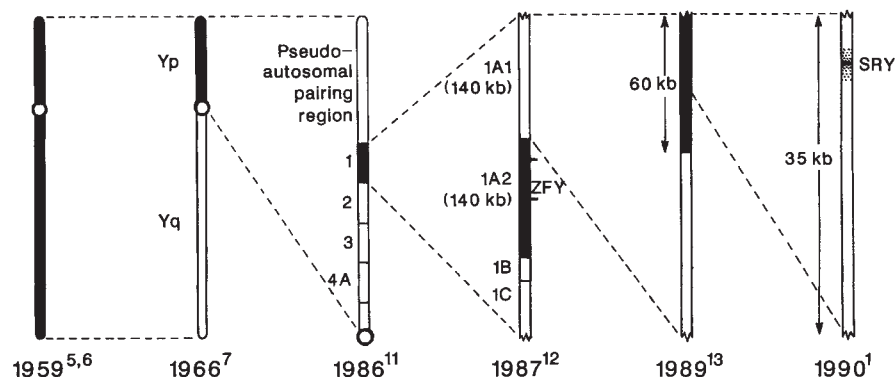
In 1959, the Y chromosome was for the first time shown to be male-determining in both mouse⁴ and man^{5,6}. Seven years later, the search in man was narrowed down to the short arm of the Y chromosome⁷. The discovery that a highly conserved, repetitive DNA sequence, sex-specific in snakes, was concentrated on the mouse Y chromosome⁸ led to much speculation as to its possible role in sex determination. But this sequence proved to be barely represented on the human Y chromosome, and interest lapsed. For nearly ten years a hypothesis postulating that the male-specific histocompatibility antigen H-Y was the primary testis-inducer⁹ held sway. In mammals, once the gonad has been committed to form a testis, maleness follows. But mice were found that had testes but lacked the H-Y antigen¹⁰, and the hypothesis was abandoned.

Next came studies of XX men who were male because one of their X chromosomes carried testis-determining sequences passed across from their father's Y chromosome. Deletion analysis suggested that the testis-determining factor (*TDF*) must be in interval 1 of the Y, adjacent to the X-Y pairing-and-exchange region¹¹ (see figure). Page *et al.*¹² then zoomed in still closer, thanks to an XX male patient whose only Y-chromosome sequences outside the pairing-and-exchange region comprised intervals 1A1 and 1A2, and a female patient with a translocation between the Y chromosome and an autosome, who appeared to lack only 1A2 and 1B. Clearly, *TDF* must be in 1A2: all 140 kilobases of DNA were isolated and cloned, and a highly conserved gene was duly located, coding for a zinc-finger-containing protein that could well function as a DNA-binding transcription regulator. *ZFY*, as the new gene was termed, seemed an excellent candidate for *TDF*. But once again, the candidate failed to be elected. Three XX men and one XX intersex who lacked *ZFY* were found¹³. All three men, however, tested positive for markers in segment 1A1. Although these men showed abnormalities of the reproductive system, suggesting that they lacked genes on the Y chromosome that are required for completely normal sexual development, all were indubitably male. This implied both that *ZFY* is not testis-

determining, and that the testis-determining gene is in 1A1. In addition, expression of the *ZFY* homologue in mice proved to be associated with the germ cells, which are believed to play no part in testis determination, rather than with the crucial somatic cell lineages of the fetal gonad¹⁴.

If the evidence from these XX men was to be believed, only 60 kilobases of segment 1A1 were left to search: this was reduced to 35 kilobases as more DNA markers became available. Progress was hampered by the abundance of repetitive sequences, but at last a single-copy gene was found¹. The amino-acid sequence

with the human sequence over 237 base pairs, and again shows strong similarities at the amino-acid level to the presumed DNA-binding motif of 80 amino acids from fission yeast. Four more such genes, showing close homology but not located on the Y chromosome, have also been detected in mice, suggesting the existence of a gene family sharing the 80-amino-acid motif. Motifs, like surnames, are relatively few in number, so it is not too surprising to find the family to which they belong scattered over the mouse genome, doing many different jobs. The four autosomal genes are all expressed in the embryo 8.5 days *post coitum*: the gene on the Y (*Sry* in the mouse) is expressed in the adult testis, and in the urogenital ridge at 11.5 days *post coitum*, the period immediately preceding sexual differentiation. Thus, the time and place of



Thirty-one years of hunting the testis-determining factor. The chromosomal region thought to include the elusive factor is shaded. The search has narrowed down from 30–40 million bases (1959) to less than 250 bases encoding the conserved 80-amino-acid motif of *SRY* (1990).

that it codes for is highly conserved, and shows homologies both with the mating-type protein Mc required for mating in fission yeast, and with the nonhistone nuclear HMG proteins thought to function as DNA-binding transcription factors. The gene has been called *SRY* (sex-determining region of the Y).

This leaves an apparent conflict with the data of Page *et al.* on the woman with the Y-autosome translocation¹². If, as claimed, she possessed the whole of 1A1, and if *TDF* lies within 1A1 (ref. 13), why was she not a man? The paradox has however been resolved by the finding that her Y-chromosome DNA has suffered a second deletion, perhaps occurring at the time of the translocation and removing just that part of 1A1 that is detected in the XX men, together with a considerable portion of the adjacent pairing-and-exchange region³.

If *SRY* is the testis-determining gene, then it ought to be expressed in the embryonic gonad at the time of its differentiation. This prediction has been tested in mice by Gubbay *et al.*, who have cloned the DNA sequence corresponding to the human *SRY* gene from the testis-determining region of the mouse Y chromosome². It shows 80 per cent homology

expression of *Sry* are consistent with a role in testis determination. But the most telling evidence comes from a mouse Y chromosome carrying a mutation that incapacitates the testis-determining gene¹⁵, producing XY females. DNA from normal XY male mice probed for *Sry* shows a 3.5-kilobase male-specific band. DNA from XY females carrying the mutant Y lacks this band, and shows no additional band, so presumably the sequences hybridizing to the *Sry* probe have been deleted to produce the mutation.

None of this evidence establishes conclusively that *SRY* (*Sry*) is *TDF* (*Tdy*). Until the deletion in the mutant Y chromosome has been better characterized, the possibility remains that *Tdy* and *Sry* are two separate but neighbouring genes, so that both have been eliminated by the same deletion. Although Sinclair *et al.*¹ have detected no other open reading frame in the 35-kilobase testis-determining region of the human Y chromosome (see figure), they admit that they could have failed to detect small exons, especially if these were closely apposed to repetitive elements.

Conclusive proof is more likely to come from mouse than from human studies. It would be reassuring if the

expression of *Sry* in the fetal testis proved to be confined to the somatic supporting cell lineage, which is believed to be critical for testis determination, rather than to the germ cells. A further test would be to insert *Sry* as a transgene into the genome of an XX mouse: a negative result would not disprove the hypothesis, because other genes on the Y chromosome might be involved, or the transgene might be expressed too late to be effective¹⁶. But if such a mouse developed as a male, it would prove that *Sry* is indeed the testis-determining gene. A genocopy of the *Tdy* mutant, knocking out *Sry* by homologous recombination to give XY females, would also be evidence hard to argue against.

But for the present, *SRY* (*Sry*) must remain no more than the most promising candidate gene yet for *TDF* (*Tdy*). If proof comes, is that the end of the matter? Should we all down tools and go home? Far from it: especially if *TDF* encodes a DNA-binding protein, we may be at the beginning of a gene-regulatory cascade at least as complex and as fascinating as those that are being unravelled in *Drosophila* and *Caenorhabditis*¹⁷. In the end we may actually understand how sex

is determined in mammals.

Afterthought: the Y chromosome is the sole genetic contribution that fathers pass on only to their sons. The present prime candidate for the human Y-borne testis-determining gene shows strong homology to the yeast gene encoding the mating type protein Mc. It is an odd coincidence that "Mc" in Gaelic denotes "son of". □

Anne McLaren is in the MRC Mammalian Development Unit, Wolfson House, University College London, London NW1 2HE, UK.

1. Sinclair, A.H. *et al.* *Nature* **346**, 240–244 (1990).
2. Gubbay, J. *et al.* *Nature* **346**, 245–250 (1990).
3. Page, D.C. *et al.* *Nature* **346**, 279–281 (1990).
4. Welshons, W.J. & Russell, L.B. *Proc. natn. Acad. Sci. U.S.A.* **45**, 560–566 (1950).
5. Jacobs, P.S. & Strong, J.A. *Nature* **183**, 302–303 (1959).
6. Ford, C.E. *et al.* *Lancet* **i**, 711–713 (1959).
7. Jacobs, P.A. & Ross, A. *Nature* **210**, 352–354 (1966).
8. Jones, K.W. & Singh, L. *Hum. Genet.* **58**, 46–53 (1981).
9. Wachtel, S.S. *et al.* *Nature* **257**, 235–236 (1975).
10. McLaren, A. *et al.* *Nature* **312**, 552–555 (1984).
11. Page, D.C. *Cold Spring Harb. Symp. quant. Biol.* **51**, 229–235 (1986).
12. Page, D.C. *et al.* *Cell* **51**, 1091–1104 (1987).
13. Palmer, M.S. *et al.* *Nature* **342**, 937–939 (1990).
14. Koopman, P. *et al.* *Nature* **342**, 940–942 (1990).
15. Lovell-Badge, R.H. & Robertson, E. *Development* (in the press).
16. Burgoyne, P.S. *Nature* **342**, 860–862 (1989).
17. Hodgkin, J. *Nature* **344**, 721–728 (1990).

NEUROPSYCHOLOGY

A nommocnu impairment

Stuart Sutherland

A CURIOUS neurological disorder, called unilateral neglect, was first noticed towards the beginning of this century: in 1941 the syndrome was more fully delineated by an English neurologist, aptly named R. Brain. People suffering from unilateral neglect have difficulty in responding to stimuli on the opposite side to their brain lesion and may fail to manipulate objects on that side of the body, for example by pulling on only one trouser leg.

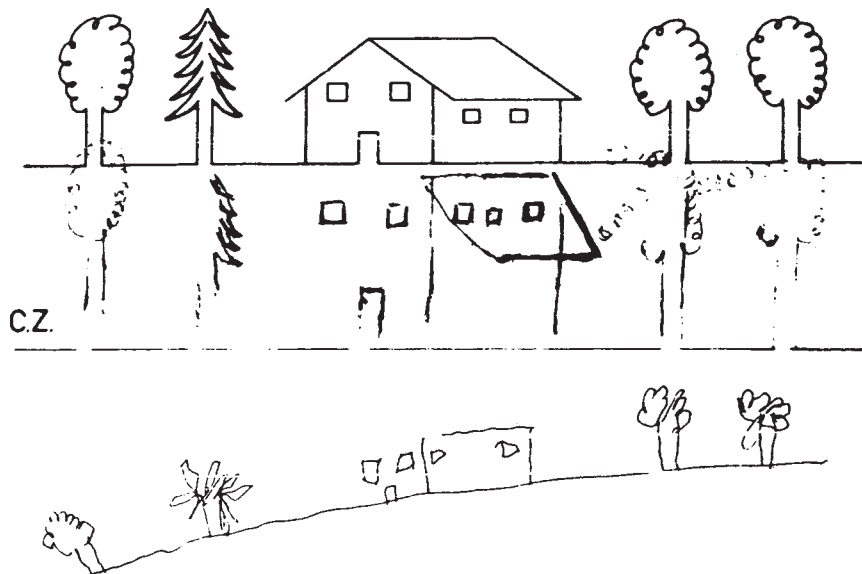
In this issue (*Nature* **346**, 267–269; 1990), Alfonso Caramazza and Argye Hillis report a new impairment that is unexpected and potentially informative. They investigated a patient with unilateral neglect caused by a lesion in the left parietal lobe; they found that she could read the left sides of words well but made gross errors with right halves. This is a not uncommon defect in such patients. With considerable ingenuity, however, Caramazza and Hillis presented her with words in other forms, in particular with words spelled backwards and with words in which the letters were arranged in a vertical string. In both cases she had problems with the second half of the words as normally written. For example, if shown 'common' spelt backwards as 'nommoc' she would read out 'com' but would make errors on 'mon' even though as far as the

eye is concerned the 'mon' (transformed into 'nom') was physically presented on the side (her left) that should not have been subject to neglect. Similarly, when words were shown vertically she still had difficulty with their second halves although there was no difference in the

position of any of the letters along the horizontal axis.

In a further series of tests, using among other tasks forward and backward spelling, the authors obtained exactly the same results. The patient performed almost perfectly on the first half of each word but made a large number of errors on the second half (over 50 per cent on the final letter). They concluded that the centre of a word must somehow occupy a fixed point in the recognition and production systems for written words and conducted a further very clever test of this hypothesis. If it is correct, the subject should perform better at recognizing 'contrast' in the word 'contrastiveness' than when presented simply with 'contrast'. Amazingly, this is exactly what happened (it would incidentally be nice to know what would happen if the additional letters were a random string so that no word is completed).

Caramazza and Hillis rightly insist that their results show that unilateral neglect can operate at a central level far removed from the topographical layout of the scene on the retina and in the early visual centres. To recognize a word, one must access its representation in memory: the results suggest that this representation preserves the left-to-right ordering of the letters in words, as indeed it must. What is puzzling is that they also suggest that words are positioned in this representation by their centre point. It is well known that words are accessed by their initial letter. When asked, most people say that more words begin with 'r' than have 'r' as the third letter but they are wrong: because we address words by their initial letter, it is difficult to retrieve words, such as 'street' that have an 'r' as their third letter. Moreover, in reading, words are usually fixated just to the right of their



Examples of copies made by unilateral-neglect patients of the composite figure shown at the top. In each case patients left unfinished half of one or more elements (see text). (Reproduced from *Neuropsychology of Visual Perception* by G. Gainotti *et al.*, courtesy of Lawrence Erlbaum.)

beginning. These facts might explain why the left-hand side of a word is easier to recognize than the right, but they do not begin to explain why a given letter string is easier to recognize if it has extra letters tacked onto the end.

Some earlier work, not mentioned by the authors, is however relevant. In their chapter in *Neuropsychology of Visual Perception* (Lawrence Erlbaum, 1989), G. Gainotti *et al.* described how they asked unilateral-neglect patients to copy a drawing of a series of objects, such as trees and houses, arranged in a horizontal row. Some of the patients copied all the objects, not just those on the same side as the lesion. But they tended to miss out those parts of each object that lay on the opposite side to the lesion (see figure). Only patients with right-sided lesions did this. It would be interesting to test Gainotti's patients with upside-down objects to see whether they would neglect the half presented contralateral to the lesion or the half of the object that would be contralateral to the lesion if the object

were in its normal orientation (as Caramazza and Hillis's patients did with words).

One possibility is that just as the left hemisphere is dominant for words, the right is dominant for the representation of objects, and that attention to the side of the representation of a word or object that is contralateral to the dominant hemisphere is impaired if input from the non-dominant hemisphere is faulty as a result of a lesion. Even this explanation, like most of those put forward by neuropsychologists, remains vague.

Nevertheless neuropsychology is coming of age: whereas in the past it concerned itself with *where* a function is carried out, it is now investigating *how* it is carried out. Its findings must be taken into account by those trying to explain how the mind works, but unfortunately many results, such as those discussed here, remain baffling. □

Stuart Sutherland is at the Centre for Research on Perception and Cognition, University of Sussex, Brighton BN1 9QG, UK.

PLATE TECTONICS

New plates, rates and dates

Jean-Bernard Minster

In the late 1960s, McKenzie and Parker¹ and Morgan² advanced the theory of plate tectonics, which proposes that the Earth's surface is covered by a mosaic of large plates that behave rigidly over geological time and move relative to each other in response to convection currents in the Earth's mantle. This theory quickly gained acceptance, particularly amongst geophysicists, and in 1968 Le Pichon put together the first self-consistent model to describe the relative motions of the major plates³. Second-generation models soon followed^{4,5}, and third-generation models were published a decade later, in 1978^{6,7}. DeMets *et al.*⁸ have now produced a revised model of plate motions that is sure to become a new standard and will probably remain so for years to come. This fourth-generation model, labelled NUVEL-1, crowns nearly a decade of effort by R. Gordon, S. Stein and their students, with the aim of improving our knowledge of plate kinematics in various regions of the world. NUVEL-1 describes the 'instantaneous' relative motions — that is, averaged over the geologically brief interval of the past 3 million years — of 12 major plates, and remedies most of the problems identified over the years with earlier models.

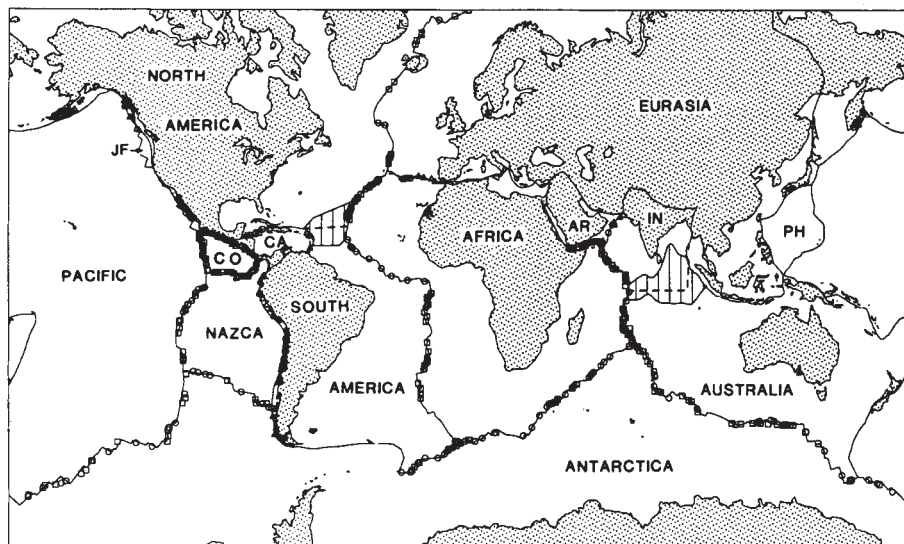
NUVEL-1 is derived from three basic data types: 277 spreading rates of new oceanic crust created at mid-ocean ridges (estimated from magnetic anomalies), 121 directions of relative plate motion (obtained from the azimuths of transform

faults, which offset the ridge crest in many places) and 724 earthquake slip vectors: a total of 1,122 data points distributed along 22 of the 30 plate boundaries in the model. This represents more than a threefold increase in sampling over earlier data sets. (As in previous studies, data from continental plate boundaries are not used because of the geological complexities associated with deformation of the weak continental crust.) NUVEL-1 accommodates all the data extremely

well, and is consistent with the hypothesis that major plates behave rigidly on a million-year timescale.

The differences between NUVEL-1 and previous models are significant. Whereas India and Australia have traditionally been assumed to belong to the same large plate, DeMets *et al.* use two plates, separated by an east-west equatorial zone of diffuse deformation (see figure). The NUVEL-1 relative plate velocities are generally slower, by about 10–15 per cent, than previous estimates. In particular, thorough reanalysis of the magnetic anomalies along the East Pacific Rise has led to spreading-rate estimates nearly 25 per cent slower than previous ones. This has important consequences. For example, consider the San Andreas fault in California: geological and geodetic observations show that it slips at a rate of 34 mm yr⁻¹ in central California, but the rigid-plate models in use until recently called for 58 mm yr⁻¹ of relative motion between the North American and Pacific plates. This so-called 'San Andreas discrepancy', which has puzzled geologists for nearly 20 years, is a considerably less severe problem with NUVEL-1, which predicts a plate slip rate of just 48 mm yr⁻¹. In particular, the apparent need for significant slip on faults other than the San Andreas along the central California margin⁹ has been lessened substantially. The concomitant reduction in estimated seismic hazard, along a coastline that harbours such sensitive facilities as the Vandenberg Space Shuttle Launch Facility and the Diablo Canyon nuclear power plant, is of no small import.

By any measure, plate tectonics has been a tremendously successful theory, and almost all Earth scientists now depend on the availability of a reliable



Data locations and plate geometry for the NUVEL-1 model⁸. Regions marked with vertical lines indicate diffuse plate boundaries (between North and South America and between India and Australia). Abbreviations of plate names are: Cocos (CO), Caribbean (CA), India (IN), Arabia (AR), Philippine (PH) and Juan de Fuca (JF).

plate-motion model. For instance, geologists need such models to provide kinematic boundary conditions in studies of continental deformation zones such as Tibet or the Basin and Range. Geodynamicists use plate velocities as boundary conditions in calculations of mantle convection.

But the group who might benefit most directly in the short term from an improved plate-motion model may be the geodetic community. Space-geodetic techniques such as very-long-baseline interferometry and satellite laser ranging can be used to measure the distance between America and Europe to within 10 mm or so, variations in the length of day to 0.05 millisecond, and the direction of the wobbling axis of rotation of the Earth to 1 milliarcsecond (that is, the position of the pole to within 30 mm); but to do so, one must account for long-term tectonic motions¹⁰. Explicit corrections must be made for plate motions, and departures from the 3-million-year average motions show up in the data. One of the exciting achievements of the past decade is the demonstration that plate motions can be measured directly by space geodesy over intervals of a few years, and that these 'current' plate motions agree very well with geological ones. In fact, in a few cases such as the western United States where abundant data are available, the 5-year space-geodetic average is in better agreement with NUVEL-1 than with previous models.

Finally, on a tectonically active planet such as the Earth, where every point of the surface is in constant motion, the very choice of a convention to define a planetary coordinate system that is stable to the necessary degree of accuracy is not possible without selecting a certain plate-motion model. This plate model then becomes an intrinsic part of the definition of the planetary coordinate system. DeMets *et al.* have provided such a model, and since the enormous amount of work that they have done is unlikely to be repeated in the near future, NUVEL-1 is a good candidate for a plate-motion standard. □

Jean-Bernard Minster is at the Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0225, USA.

1. McKenzie, D.P. & Parker, R.L. *Nature* **216**, 1276–1280 (1967).
2. Morgan, J.W. *J. geophys. Res.* **73**, 1959–1982 (1968).
3. Le Pichon, X. *J. geophys. Res.* **73**, 3661–3697 (1968).
4. Chase, C.G. *Geophys. J. R. astr. Soc.* **29**, 117 (1972).
5. Minster, J.B., Jordan, T.H., Molnar, P. & Haines, E. *Geophys. J. R. astr. Soc.* **36**, 541–576 (1974).
6. Chase, C.G. *Earth planet. Sci. Lett.* **37**, 353–368 (1978).
7. Minster, J.B. & Jordan, T.H. *J. geophys. Res.* **83**, 5331–5354 (1978).
8. DeMets, C., Gordon, R.G., Argus, D.F. & Stein, S. *Geophys. J.* **101**, 425–478 (1990).
9. Minster, J.B. & Jordan, T.H. *J. geophys. Res.* **92**, 4798–4804 (1987).
10. Muller, I.I. & Zerbini, S. *The Interdisciplinary Role of Space Geodesy* (Springer, Berlin, 1989).

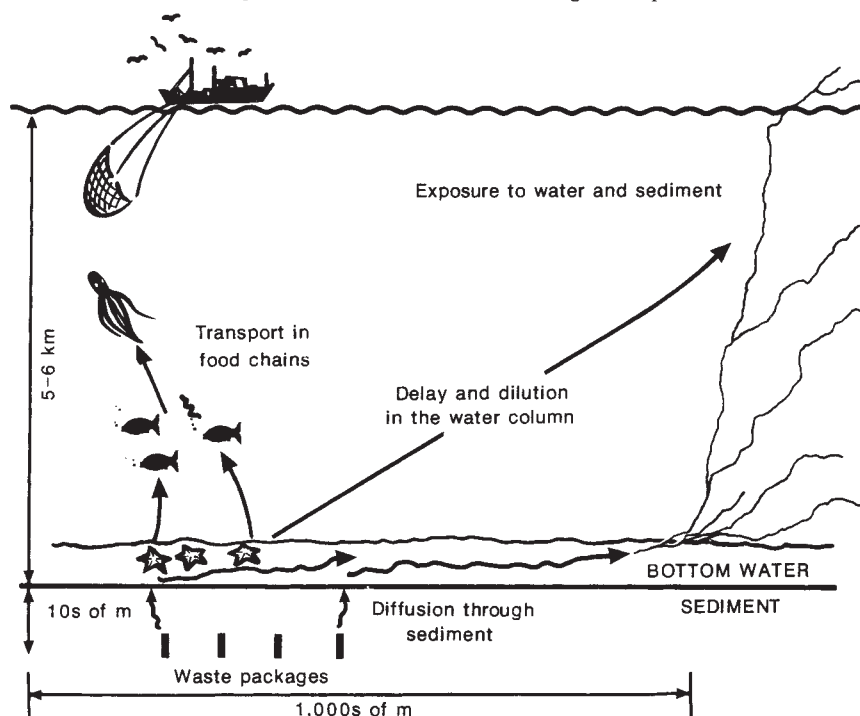
Lessons from the deep sea

J. Kirk Cochran

THE safe disposal of high-level radioactive waste generated by nuclear power plants, fuel reprocessing facilities and the manufacture of atomic weapons remains a serious problem. Although most of the disposal methods proposed are land based, an alternative is burial within sediments of the deep ocean (see figure). Sub-seabed disposal relies on the existence of several barriers that retard the spread of the radioactivity and minimize contamination of food webs¹. Vitrification of the waste and its packaging in canisters constitutes the first barrier to migration of radionuclides, containing the waste for the 10–100 years during which it is at a

how natural experiments can be exploited to assist in the evaluation of the deep sea as a nuclear-waste repository.

The great abyssal plains of the North-east and Northwest Atlantic Ocean, at depths of 5,000–6,000 m, receive periodic inputs of sediment carried down the continental slope by turbidity flows. These turbidites comprise mostly organic-rich detrital sediment, in contrast to pelagic sediments which comprise the remains of planktonic organisms that settle slowly but continuously to the sea floor. The organic-rich nature of the turbidites and the geochemical conditions characteristic of their original depositional environment



Sub-seabed disposal of high-level nuclear waste. From ref. 4.

high temperature because of the heat generated by radioactive decay. Over the longer term (100–100,000 years), the few tens of metres of sediments between the waste container and the overlying water form a second barrier.

On page 260 of this issue², Colley and Thomson describe a means of testing the efficacy of the sediment barrier over sediment depths and timescales appropriate for the containment of radioactive waste. By evaluating the relationship between the distribution of natural ²³⁸U and the products of its radioactive decay in turbidite sediments, they show that the rate of migration of these radionuclides is negligible (except for that of ²²⁶Ra) and that, for all the radionuclides considered, there has been no loss from the two turbidite sections studied into the overlying sediments or the bottom water. This study illustrates

make them rich in naturally occurring uranium. The turbidites are deposited initially with a relatively constant concentration of uranium throughout their thickness. As the top of the turbidite is oxidized through contact with dissolved oxygen in the bottom water, the uranium that it contains becomes mobilized and migrates downwards.

But when the uranium reaches the anoxic sediments deeper in the turbidite, it is immobilized, and thus a sharp concentration gradient is established. This process of uranium redistribution stops when contact with oxygenated bottom water is broken through the deposition either of sufficient pelagic sediment or of another turbidite. Turbidite units buried deeper in the sediment column therefore have increasing concentrations of radioactive daughters produced by decay of uranium.

Given sufficient time (more than 500,000 years) and no migration, the radioactivities of each member of the decay chain should be equal, a condition known as secular equilibrium. The stratigraphy of the Madeira Abyssal Plain has been sufficiently well established¹ for Colley and Thomson to select two turbidite units older than 500,000 years (buried about 25 metres below overlying sediment) from a 30-m piston core taken there. They examined several parent/daughter pairs in the ²³⁸U decay series to look for evidence of migration: ²³⁸U/²³⁴U, ²³⁴U/²³⁰Th, ²³⁰Th/²²⁶Ra and ²²⁶Ra/²¹⁰Po (²¹⁰Pb).

These comparisons enable the authors to assess the relative mobility of the elements U, Th, Ra and Po (Pb) deep in the sediment column over timescales of about 500,000 years. Colley and Thomson find that uranium and thorium have not migrated significantly, but that radium has migrated over distances of about 20 cm. For the turbidite units as a whole, ²³⁸U was in secular equilibrium with its daughters, including ²²⁶Ra.

These results are important because they provide a direct test of long-term radionuclide migration at depth in the sediment column under *in situ* geochemical conditions. Moreover, many of the

actinides produced artificially in high-level radioactive waste form decay chains containing isotopes of the same element as are present in the natural ²³⁸U, ²³⁵U and ²³²Th decay series. The approach of Colley and Thomson cannot, however, take into account the advection of pore water likely to result from the high temperatures near the freshly emplaced waste canister (the 'near-field') and is thus more appropriate for evaluations of radionuclide migration in the 'far-field'. Studies such as the In Situ Heat Transfer Experiment were planned by the US Subseabed Disposal Program to investigate near-field transport, but were cancelled owing to lack of funding. But although land-based disposal of radioactive waste seems to be the approach most likely to be adopted, we should not abandon all efforts to evaluate the deep sea as an alternative repository. □

J. Kirk Cochran is at the Marine Sciences Research Center, State University of New York, Stony Brook, New York 11794-5000, USA.

1. Hinga, K.R., Heath, G.R., Anderson, D.R. & Hollister, C.D. *Envir. Sci. Technol.* **16**, 28A–37A (1982).
2. Colley, S. & Thomson, J. *Nature* **346**, 260–263 (1990).
3. Weaver, P.P.E. & Rothwell, R.G. in *Geology and Geochemistry of Abyssal Plains* (eds Weaver, P.P.E. & Thomson, J.) 71–86 (Geological Society of America, 1987).
4. Kaplan, M.F., Koplik, C.M. & Klett, R.D. *Sandia Natl. Lab. Rep.* SAND83-7106 (1984).

LOCOMOTION

Running is priced by the step

R. McNeill Alexander and R.F. Ker

AT last we have a good explanation for why when running, small mammals proportionately use so much more energy than large ones. Measurements of rates of oxygen consumption reported by Kram and Taylor on page 265 of this issue¹ show that 140-kilogram ponies use 0.2 joules of metabolic energy per newton of body weight, for every metre that they run, at any speed. In contrast, 30-gram kangaroo rats use 2 joules per newton metre. Kram and Taylor's explanation for the discrepancy arises from their measurements of the time feet spend in contact with the ground, and also from measurements² of the mechanical advantages of limb muscles. Smaller mammals have shorter foot-ground contact times than large ones, so they need faster muscles that are less economical of energy: also, these muscles have to exert larger forces relative to body weight.

At some stages of the stride, active muscles shorten and do work. At others, muscles are forcibly extended and do 'negative work', degrading mechanical energy to heat. The negative work cancels out nearly all the positive work when the animal is running at constant speed over level ground because very little net work — merely enough to overcome air resistance and joint friction³ — is then needed.

Whenever muscles exert force, their crossbridges cycle and metabolic energy is used. Energy is used faster (for the same force) when the muscle is shortening, and slower when the muscle is being stretched^{4,5}, but the extra cost of the work done at some stages of the stride is probably largely balanced by savings due to negative work at others. So to estimate the energy cost of running we need consider only the cost of exerting force.

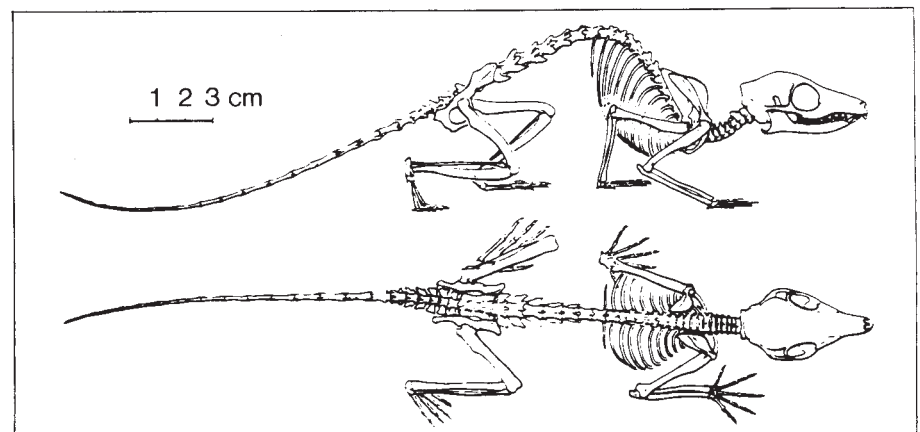
The metabolic energy used in running

should be proportional to the number of crossbridge cycles (the number of crossbridges multiplied by the number of cycles made by each). If foot-ground contact time is made shorter, faster muscles with higher crossbridge cycling rates must be used. Thus the rate of use of metabolic energy should be proportional to the volume of muscle, multiplied by the stress it must develop (averaged over a stride), divided by contact time.

Mammals of different sizes have about the same proportion of leg muscle in their bodies⁶, and during locomotion the same stresses act in these muscles. The rate of use of metabolic energy per unit weight should therefore be inversely proportional to contact time, which Kram and Taylor show to be the case. In other words, mammals of all sizes use about the same amount of energy moving unit weight through one step length (the distance travelled during one foot contact time).

A mammal's step length is about the same at all speeds, but is limited by leg length, and Kram and Taylor find it to be proportional to (body mass)^{0.30}. Thus the energy cost of moving unit weight of animal unit distance should be proportional to (body mass)^{-0.30}. It is actually very close to this, about proportional to (body mass)^{-0.25}.

The key to Kram and Taylor's story is that mean leg muscle stresses are about the same in mammals of all sizes². If animals were geometrically similar, the forces that their muscles would have to exert to support their body weights would be proportional to body mass, but the cross-sectional areas of the muscles would be proportional to (body mass)^{0.67}. Muscle stresses would be proportional to (body mass)^{0.33}, about 15 times higher in 300-kilogram horses than in 90-gram chipmunks. They are actually about the same, because chipmunks (and other small mammals; see figure) run in a crouched posture with knees and elbows bent, making their muscles work with very low mechanical advantages. Horses (and



Running on bent legs, as this tree shrew does, is expensive of energy, but small mammals may adopt this gait so that they can accelerate or jump quickly. This outline, by F.A. Jenkins⁷, is based on X-ray cinematography.

other large mammals) keep their legs much straighter.

Why do chipmunks not run on straighter legs, like tiny horses, and save a large proportion of the energy needed for running? One possible answer is that, with bent legs, they are immediately ready either to accelerate or to jump. Large mammals (including ourselves) have to bend their legs before jumping. It seems likely that evolution may favour some compromise between readiness to take evasive action, or to pounce, and economy.

We assume that economy is important even though it may not be the sole determinant of optimal gait. If the assumption is correct, Kram and Taylor's report may explain why antelopes have long legs and cheetahs have flexible backs. These have long been regarded as adaptations that increase speed by lengthening the stride,

but previous theories of the energetics of running made the interpretation seem doubtful because longer strides need more work. The new theory shows how the longer steps that long legs and flexible backs make possible may reduce the metabolic cost of running at a given speed. □

R. McNeill Alexander and R.F. Ker are in the Department of Pure and Applied Biology, University of Leeds, Leeds LS2 9JT, UK.

1. Kram, R. & Taylor, C.R. *Nature* **346**, 265–267 (1990).
2. Biewener, A.A. *Science* **245**, 45–48 (1989).
3. Alexander, R.McN. in *Mechanics and Energetics of Animal Locomotion* (eds Alexander, R. McN. & Goldspink, G.) 168–203 (Chapman & Hall, London, 1977).
4. Woledge, R.C., Curtin, N.A. & Homsher, E. *Energetic Aspects of Muscle Contraction* (Academic, London, 1985).
5. Heglund, N.C. & Cavagna, G.A. *Am. J. Physiol.* **253**, C22–C29 (1987).
6. Alexander, R.McN., Jayes, A.S., Maloiy, G.M.O. & Wathuta, E.M. *J. Zool.* **194**, 539–552 (1981).
7. Jenkins, F.A. in *Primate Locomotion* (ed. Jenkins, F.A.) (Academic, New York, 1974).

NEUTRINO PHYSICS

The revenge of orthodoxy

Pierre Salati

How many neutrino families are there? What are the cosmological consequences that may be derived from the large number of Z^0 decay events collected at the CERN Large Electron–Positron (LEP) collider? Is the flux of neutrinos from the solar core in agreement with the predictions of the Standard Model? Is this flux related to sunspot activity? Could the electron neutrino supply enough dark matter to close the Universe? These and other questions show that neutrino physics is at the frontier between astrophysics, cosmology and particle physics. Researchers from these various fields gathered recently at CERN to discuss the latest observations relating to such issues. One conclusion was that the Standard Model of electroweak interactions has survived well the impressive tests conducted at LEP over the past year, which imply that there are just three neutrino species. And the deficiency of solar neutrinos relative to the predicted flux has been confirmed by the latest results from the Kamiokande II neutrino detector which, however, finds no correlation between the data and the sunspot cycle.

The four LEP detectors (ALEPH, DELPHI, L3 and OPAL) have collected more than 150,000 Z^0 decay events over its first year of operation. The neutral Z^0 , which mediates weak interactions, may decay into quarks, charged leptons (such as the electron and the muon) and neutrinos. Each neutrino species contributes 167 MeV to the total decay width of the Z^0 , which is determined most accurately from the cross-section at the resonance

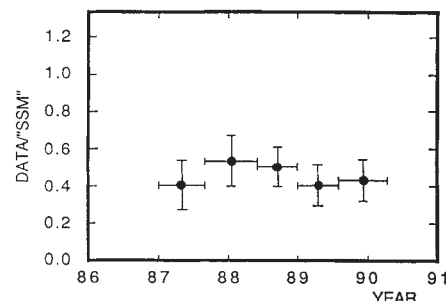
peak. Hadrons (quarks) and charged leptons produced in the decays are observed directly, so their contributions to the decay width are readily measured. The neutrino contribution (Γ_ν) may thus be deduced; the result, $\Gamma_\nu = 496 \pm 18$ MeV, puts the number of neutrino species N_ν at 2.99 ± 0.09 (see table). The Standard Model is vindicated in this respect (U. Becker, MIT); three neutrino families are already known to exist and the general assumption is that $N_\nu = 3$. A fourth-generation neutrino is forbidden by the LEP results unless its mass exceeds 45 GeV. If this is so, the relic abundance of such a heavy neutrino — which has been one of the most favoured candidates for astronomical dark matter — is too low to be of cosmological significance¹.

NEUTRINO CONTRIBUTION TO Z^0 DECAY WIDTH

Experiment	$\Gamma_\nu/3$
ALEPH	164 ± 8
DELPHI	171 ± 15
L3	172 ± 14
OPAL	158 ± 14
Stanford Linear Collider	153 ± 33
Theory	167 ± 1

Latest measurements of the solar neutrino flux were reported from the Japanese neutrino detector Kamiokande II (see figure). Neutrinos scatter from the electrons of water molecules in the detector tank, causing the electrons to emit Cerenkov radiation. Detection of this radiation gives information about both the direction and the energy of the incoming neutrinos. The flux of those originating in the Sun is found to be lower by a factor of two than the predictions of standard solar models (Y. Totzuka, Institute for Cosmic Ray Research,

Tokyo). The lack of evidence for a correlation between the sunspot cycle and the ^8B neutrino flux (that is, those from solar nuclear reactions involving ^8B , which have energies ≥ 1 MeV) in the Kamiokande data contrasts with the results from the Homestake detector in South Dakota (K. Lande, University of Pennsylvania),



The ratio of measured (by Kamiokande II) to predicted solar neutrino flux as a function of time. Data have been accumulated for 1,040 effective days since 1987. The ratio is systematically lower than unity and, for $E_\nu > 7.5$ MeV, is equal to 0.46 ± 0.05 (statistical) ± 0.06 (systematic), in agreement with results from the Homestake detector.

which show a significant drop in the argon count rate (produced from chlorine in the tank through the reaction $^{37}\text{Cl} + \nu_e \rightarrow ^{37}\text{Ar} + e^-$) as solar activity approaches its maximum. Are the Homestake data statistically significant, asked J. Bahcall (Institute for Advanced Study, Princeton)? If so, a possible explanation^{2,3} requires an electric or magnetic dipole moment for the electron neutrino of the order of 10^{-11} Bohr magnetons (μ_B). With a moment of this magnitude, left-handed neutrinos produced at the solar centre would be flipped to right-handed helicity while crossing the convective layer close to the solar surface, inside which magnetic fields are suspected to exist, and would therefore become invisible because right-handed neutrinos do not interact with matter. But a fairly strong magnetic field is needed to achieve this; and furthermore, an upper bound of $3 \times 10^{12} \mu_B$ on the electron–neutrino dipole moment may be inferred from the evolution of low-mass stars along the red-giant and subsequent horizontal branches of the Hertzsprung–Russell diagram (G. Raffelt, Max-Planck-Institut für Physik, Munich), making this scheme rather contrived. Surprisingly, Kamiokande finds on average a larger flux (3.6 solar neutrino units) than Homestake (2.3 SNU). But as the energy threshold for neutrino detection is larger for the former (7.5 MeV) than for the chlorine–argon transmutation reaction of the latter, a central solar temperature lower than normal — another proposed solution of the solar neutrino deficiency — would give precisely the opposite result, producing more low-energy neutrinos which Kamiokande would fail to detect.

The gallium detectors currently being

* Neutrino 90, 14th International Conference on Neutrino Physics and Astrophysics, CERN, Geneva, 10–15 June 1990.

prepared will elucidate conditions in the deep solar core. Their fairly low detection threshold of 233 keV will make them sensitive to the entire neutrino spectrum, in particular to those produced by the collision of two protons. The Soviet-American gallium experiment SAGE is already yielding interesting preliminary results (V.N. Gavrin, Institute for Nuclear Research, Moscow). From the data accumulated during the first four months of 1990, SAGE has found a proton-proton neutrino flux of 0 SNU. Errors are still fairly large, however: SAGE gives an upper limit of 138 SNU at the 95 per cent confidence level. The Standard Model prediction of 125–130 SNU will soon be tested.

The Italian-American collaboration MACRO has presented its first measurements of the muon flux (one product of interactions involving cosmic neutrinos) through its detector beneath the Gran Sasso mountain. The MACRO experiment has excluded scalar neutrinos (the spin-zero bosons that supersymmetric theories associate with ordinary neutrinos) as the dark-matter component of the Galactic halo. If scalar neutrinos pervade the Galaxy, they would accumulate steadily inside the Sun and would annihilate to produce muon neutrinos, which would be converted into a detectable muon flux when they pass through the Earth.

Two tritium β -decay experiments have yielded bounds on the electron neutrino mass. The high-energy tail of the electron spectrum produced in β decays depends fairly sensitively on the mass of the electron neutrino. From studies of a monolayer of tritium on an organic substrate, a European group (E. Holzschuh, University of Zurich) has concluded that the electron neutrino mass cannot exceed 15.4 eV. And by studying a beam of molecular tritium gas, an American team (J. Wilkerson, Los Alamos National Laboratory) has obtained an upper bound of 9.4 eV. Electron neutrinos alone do not therefore seem able to close the Universe, as they are too light.

Aside from the solar-neutrino deficiency, which remains a puzzling and exciting problem, the meeting represented the revenge of orthodoxy over unconventional ideas. For those who favour the heterodox, the status of the Standard Model is still distressingly good (G. Altarelli, CERN), and becomes increasingly so with the passage of time. $\square$

Pierre Salati is in the Theory Division, CERN, CH-1211 Geneva 23, Switzerland, and is on leave of absence from LAPP, BP110, 74941 Annecy-le-Vieux Cedex, France and the University of Chambéry, France.

Reading between the lines

Ira Wasserman

THE magnetic fields of neutron stars have long been thought to be very large, but direct evidence of this is scarce. Cyclotron lines in the X-ray and γ -ray spectra of these objects provide such a direct probe. They are known to be the result of transitions between the quantized energy levels of electrons circulating the strong fields, but their precise interpretation has been fraught with ambiguity. On page 250 of this issue¹, Mihara *et al.* report on observations of the X-ray cyclotron line of the X-ray pulsar Hercules X-1, obtained by the Ginga satellite. The high quality of these observations allows new insight into the mechanism of cyclotron-line formation, and thus into the neutron-star environment.

Simple arguments about the 'freezing' of magnetic flux demonstrate that neutron-star magnetic field strengths should be of the order of 10^{10} gauss or larger. Indirect evidence derived from studies of the spin-down of radio pulsars² and the spin-up of X-ray pulsars³ suggested that their fields are indeed very large and moreover are quite similar for the two types of object — typically around 10^{12} gauss, quite close to naive expectations.

For magnetic fields of strength $B \approx 10^{12}$ gauss, the energy separation between successive quantized orbits of electrons circling the field (the so-called Landau levels) is roughly $\hbar e B / m_e c \approx 11.6 B_{12}$ keV (where $B_{12} = B/10^{12}$ gauss, e and m_e are the electron's charge and mass, and c is the speed of light), which lies in the 'hard', or high-energy, X-ray domain. Landau orbits play a critical role in the structure of the atmospheres of accreting neutron stars; the electron plasma is essentially one-dimensional, with the electrons almost exclusively in the Landau ground state at the relevant densities and temperatures⁴. Cyclotron lines corresponding to transitions between the ground and first excited state should be detectable in the X-ray spectra of most, if not all, X-ray pulsars.

The first cyclotron line to be observed was detected in the high-energy spectrum of Her X-1 by Trumper *et al.*⁵ in 1978. Because of the difficulties involved in searching for spectral lines at unknown X-ray energies^{6,7}, it was not obvious at first whether the observed cyclotron feature was an emission line at about 58 keV or an absorption line at about 38 keV. In fact neither interpretation is strictly correct: the line is most probably a 'resonant scattering' feature at about 38 keV. At the relatively low electron densities prevalent in neutron-star atmospheres, photons with energies near that of the energy spacing between the Landau ground and first excited states are more or less never

truly absorbed, but they are shifted away from the resonance as they traverse the atmosphere, giving the superficial appearance of absorption.

Cyclotron lines have now been detected, with varying degrees of reliability, in the spectra of a number of X-ray pulsars^{8–10}. The results of Mihara *et al.*¹ represent an improvement on earlier studies of Her X-1: they show that the cyclotron line is centred at roughly 35 keV, corresponding to a magnetic field strength for Her X-1 of $B_{12} \approx 3$ (uncorrected for surface redshift). The line is rather broad ($\Delta E/E \approx 0.9$, ΔE being the full width of half maximum), roughly ten times the energy resolution of the detector near the line energy. This fairly large value of $\Delta E/E$ suggests that cyclotron-line formation is probably a relatively broad-band phenomenon in X-ray pulsars, extending over a wide energy range, as has long been suspected on theoretical grounds^{11–14}. Mihara *et al.* quantify this suspicion with a crude analytic model and argue that the exponential fall-off of the Her X-1 continuum spectrum at high energies may be due largely to absorption in the wings of the first two Landau transitions. In my view, the authors rather overstate their case when they assert that their fits confirm that the line feature is an absorption line. But the X-ray continua of accreting, magnetic neutron stars are probably altered significantly by cyclotron resonant and Raman scattering¹⁵, at least in the qualitative manner advocated by Mihara *et al.*

Unfortunately, the Ginga observations did not quite extend up to sufficiently high energies to detect the first harmonic cyclotron line, resulting from transitions from the ground to the second Landau excited state at approximately 70 keV. In their discovery paper, Trumper *et al.*⁵ claimed to have detected a spectral feature near 100 keV at low statistical significance; this putative feature was attributed to the first harmonic cyclotron transition. Subsequent observations have not corroborated that claim. Landau excited states decay primarily by electric dipole transitions in which the radiated photon has an energy given approximately by $\hbar e B / m_e c$, corresponding to the (nearly equal) spacing between adjacent Landau levels. Excitation of a ground-state electron to the N th Landau level by absorption of a single photon with energy of about $N \hbar e B / m_e c$ is usually followed by a series of N decays producing N photons near the energy of the cyclotron fundamental, $\hbar e B / m_e c$. Thus photons scattered at harmonics are destroyed, and the harmonic resonances closely resemble true absorption lines. As a result, one can deduce relatively easily

1. Griest, K. & Silk, J. *Nature* **343**, 26–27 (1990).

2. Voloshin, M.B., Vysotskii, M.I. & Okun, L.B. *Sov. J. Nucl. Phys.* **44**, 440–441 (1986).

3. Voloshin, M.B. & Vysotskii, M.I. *Sov. J. Nucl. Phys.* **44**, 544–545 (1986).

the characteristics of the electron plasma from observations of cyclotron harmonics. Pairs of cyclotron lines, corresponding to $N = 1$ and 2 Landau transitions, have so far been detected in the spectra of one X-ray pulsar, 4U0115+63⁹, and of two γ -ray bursts, GB880205 and GB870303^{7,10}. For GB880205, the higher harmonics have been of great value in determining physical conditions in the line-forming region^{7,15}.

Because the cyclotron lines are formed near or on the surfaces of neutron stars, the detection of a pair of high-energy lines is in principle sufficient to deduce, at the very least, a lower bound on the surface redshift of the line-forming region. The surface redshift imposes significant constraints on theories of nuclear matter in extremely condensed states. For Her X-1, (special) relativistic corrections to the Landau energy levels are non-negligible, implying departures from equal level spacings of approximately 10 per cent. As this relativistic shift is proportional to the magnetic field strength, one could hope to deduce B from the ratios of line-centre energies, and thus to calculate the redshift from the energies themselves. In reality,

detailed radiative transfer effects probably displace the line-centre energies in highly nonlinear ways that are dependent on the properties of the scattering atmosphere. An attempt was made to deduce the redshift of the line-forming region in GB880205 by Wang *et al.*¹⁵, but without success. □

Ira Wasserman is in the Department of Astronomy, Cornell University, Ithaca, New York 14853, USA.

1. Mihara, T. *et al.* *Nature* **346**, 250–252 (1990).
2. Gold, T. *Nature* **218**, 731–732 (1968).
3. Lamb, F.K., Pethick, C.J. & Pines, D. *Astrophys. J.* **184**, 271–289 (1973).
4. Ventura, J. *Phys. Rev.* **A8**, 3021–3038 (1973).
5. Trumper, J. *et al.* *Astrophys. J.* **219**, L105–L110 (1978).
6. Holt, S.S. & McCray, R. *A. Rev. Astr. Astrophys.* **20**, 323–365 (1982).
7. Fenimore, E.E. *et al.* *Astrophys. J.* **335**, L71–L75 (1988).
8. Wheaton, W.A. *et al.* *Nature* **282**, 240–243 (1979).
9. White, N.E., Swank, J.H. & Holt, S.S. *Astrophys. J.* **270**, 711–734 (1983).
10. Clark, G.W. *et al.* *Astrophys. J.* **353**, 274–280 (1990).
11. Pravdo, S.H. *et al.* *Astrophys. J.* **225**, 988–993 (1978).
12. Wasserman, I. & Salpeter, E.E. *Astrophys. J.* **241**, 1107–1121 (1980).
13. Meszaros, P. & Nagel, W. *Astrophys. J.* **298**, 147–160 (1985).
14. Wang, J., Wasserman, I. & Salpeter, E.E. *Astrophys. J.* **338**, 343–358 (1989).
15. Wang, J. *et al.* *Phys. Rev. Lett.* **63**, 1550–1553 (1989).
16. Murakami, T. *et al.* *Nature* **335**, 234–235 (1988).

DEVELOPMENTAL BIOLOGY

A handle on handedness

John Galloway

THE problem of how the handedness of an enantiomorphic structure is assigned by a developing organism is deeply interesting yet curiously neglected — few models of the process seem to exist. In a recent issue of *Development*, Brown and Wolpert¹ take a shot at remedying this unsatisfactory state of affairs by using the phenomenon of *situs inversus viscerum* — the occasional reversal of visceral asymmetry seen in mammals — as a hook on which to hang a theory.

At the molecular level, biological structure is decidedly and consistently biased. At higher levels we find no such consistency. Some organs, the heart for instance, and some organisms, for example the gastropods², possess a marked asymmetry and one that is under genetic control, although the nature of that control remains poorly understood. Some enantiomorphic structures occur as mutual mirror-image pairs reflected in their possessor's midplane — arms, legs and horns for instance. But many others have their handedness or screw-sense assigned to them apparently at random; phyllotaxis offers many good examples^{3,4}. *Situs inversus* is an instructive example of a halfway house — the removal of genetic control by mutation results in a random assignment of handedness.

Because the viscera are disposed asymmetrically in mammals, it is possible for

individuals to be born with *situs inversus*, that is with their normal visceral symmetry reversed. And indeed this occurs, although it is rare in human beings, affecting less than one in 10,000 people. In mice, *situs inversus* is the result of an autosomal recessive gene (designated *iv*, although it can occur spontaneously and as the result of some other mutations; see Fig. 1). In homozygotes, the ratio of normal to reversed symmetry is 1:1. This led Layton⁵ to propose that the normal allele at the *iv* locus specifies normal symmetry, its absence leading to the random allocation of 'normal' and reversed symmetries (Annett⁶ essentially proposed the same for the genetics of left-handedness).

It looks very much therefore as though two distinct processes are at work: a stochastic one that, left to its own devices, randomly assigns structural differences to the two sides of the embryonic midplane; and a deterministic one that can overbear its vacillating colleague. What could the nature of such processes be? Morphogenic gradients are strong candidates — particularly as Driever and Nüsslein-Volhard⁷ have demonstrated that the anteroposterior axis of the developing

Drosophila embryo is defined by such a gradient.

A morphogen diffusing symmetrically from the midplane provides a credible explanation for the development of bilateral (mirror) symmetry. Brown and Wolpert¹ suggest that asymmetry could be produced by diffusion-reaction, which does seem capable of subdividing an organism into separate distinct compartments and theoretically could be random. If gradients can lead to the development of distinct longitudinal segments in *Drosophila*⁸ then they might be able to develop differences in the left and right halves of an embryo.

And the deterministic mechanism? Because this is under immediate genetic control, it presumably resides in the structure of a protein molecule or an assembly of proteins. Forget for a moment the particular problem presented by *situs inversus*. Could genes directly control the handedness of some biological structures? Through controlling the symmetry of a protein, or proteins, this seems unlikely. Only L-amino acids occur in proteins and as a result nature favours (virtually exclusively) the right-handed α -helix. And indeed all structural features of proteins where there is a choice of handedness favour one hand over the other⁹.

The same considerations apply to nucleotides and DNA. Some D-amino acids do occur in nature — in dermorphin (and in other neuropeptides) and in the peptidoglycans of bacteria, and a unique reverse-turn of α -helix in thermolysin. And among the polynucleotides, Z-form DNA is the exception that disproves the rule. But despite some theoretical claims to the contrary, the odd idiosyncrasy and some explanatory difficulties, ambidex-

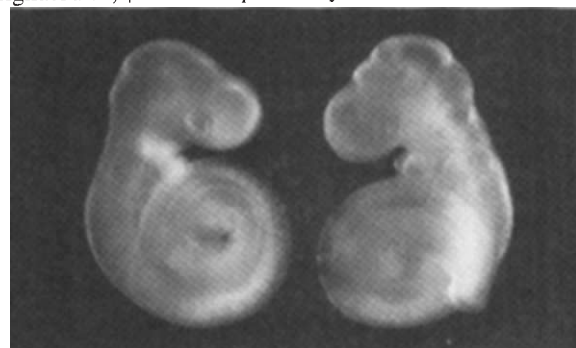


FIG. 1 Normal (left) and *situs inversus* (homozygous *iv/iv*) 10.5-day mouse embryos.

trousness is not a feature of molecular biology. It seems from what we know of the determinants of protein structure, that genetic mutation could not result in a reversal of symmetry. This firmly rules out the idea, put forward from time to time, that symmetry at the level of organ or organism could be a direct consequence of molecular symmetry.

Brown and Wolpert suggest instead that molecular symmetry may well determine handedness at a higher structural level but

point out that in an organism, left- and right-handedness is defined only with respect to the anteroposterior and dorso-ventral axes. If the origins of handedness do reside in something a protein does, then the protein must be orientated with

the other hand it does not immediately seem to offer insights into left- and right-handedness in gastropods. A recessive gene usually represents a loss of function in the normal gene product. This is one reason why Layton's suggestion is attractive (and entirely consistent with Davis's findings³ of non-genetic control of left- and right-handedness in palm trees for instance).

But for gastropods the loss of gene function does not lead to an expected and comprehensible replacement of a deterministic mechanism by a probabilistic one, but by another deterministic one. Freeman and Lundelius¹⁰ found a cytoplasmic protein that will convert the abnormal left-handed screw sense of the snail *Limnaea peregra* to the normal right-handed one. But no equivalent protein determining left-handedness is part of the normal development process which a separate gene can convert to right-handedness — there seems to be nothing probabilistic about the development of gastropods (or very little). In normally left-handed gastropods like *Partula* the whole scheme would be reversed of course.

Whether or not the approach of Brown and Wolpert is attractive is, for now, a matter of taste. What is important is that they have suggested a philosophically sound scheme, very much in the spirit of modern embryology, of which some aspects could be tested — at least in principle. Some progress has already been made, for instance in mapping the *iv* gene¹¹, so it is on the cards

that the nature of the protein will soon be known. □

John Galloway is at the Cancer Research Campaign, 2 Carlton House Terrace, London SW1Y 5AR, UK.

Up the carbon path

OVER half the take-off weight of a conventional space rocket is oxygen. Hence the interest in air-breathing space planes, rockets launched from aircraft to minimize first-stage weight, electromagnetic rail-launchers, even Iraqi-type super guns. Daedalus's contribution to the debate is a neat and novel air-breathing ramjet.

Its motor is simply a large block of graphite with a Venturi tube bored through it. The tube is initially filled by a charge of solid propellant. On igniting the rearward surface of this charge, it powers the vehicle into the air while heating the inner graphite surface white-hot. By the time the charge has burned away, allowing air to enter the Venturi, the craft is travelling fast enough to act as a ramjet. The air rammed into it by its forward motion burns the internal surface, and the resulting hot exhaust generates thrust.

A motor constructed entirely of fuel, of course, saves so much structural weight that it can accelerate a respectable payload to a very high final velocity. Even better, Daedalus's graphite ramjet is self-optimizing. As it rises through the atmosphere, the air pressure around it steadily falls. But burning steadily away from inside, the carbon Venturi automatically widens to capture more air, and maintains its thrust and efficiency. DREADCO's hydrodynamicists are working hard to define that initial shape of Venturi whose steady ablation by burning generates the ideal thrust envelope for space entry, and is totally consumed by the time the vehicle has risen out of the useful atmosphere. Unlike a rocket, the craft may have to climb at an angle to breathe air as long as possible.

Sooner or later, however, a rocket stage must take over to insert the craft in orbit. But rather than launching a separate second stage, Daedalus is designing an enlarged graphite ramjet which gradually turns into a rocket at the appropriate stage of the climb. Its outer layers are loaded with a graded proportion of a solid oxidizer like ammonium perchlorate. As ram-action begins to fail, combustion will reach these layers, thus compensating for the increasing oxygen deficiency. As the motor becomes a pure rocket, a multi-leaf one-way valve (almost the only moving part in the whole assembly) will seal the air inlet to prevent the propulsion gases emerging in the wrong direction.

This elegant and efficient space vehicle, lifting little deadweight and no more oxygen than it needs, should reduce satellite launching to a cheap and reliable routine. With its almost arbitrarily high fuel/payload ratio, it will speed small space probes right up to Earth-escape velocities as well, for interplanetary exploration.

David Jones

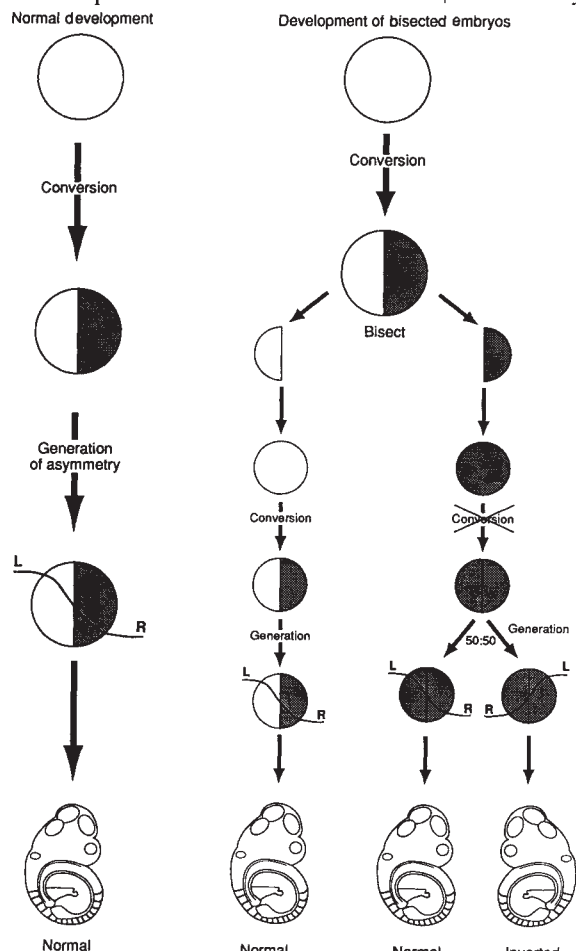


FIG. 2 An explanation of the anomalous development of hemi-embryos and conjoined twins. "Conversion" (ref. 1) is the deterministic mechanism that converts a molecular asymmetry at the cellular level and results in some stable property on the right half of an embryo. This in some way biases the otherwise random generation of asymmetry. The left half of an embryo bisected after "conversion" can develop normally, whereas the right half develops with random asymmetry because the stable property of conversion is present in both halves. (From ref. 1.)

respect to these two axes.

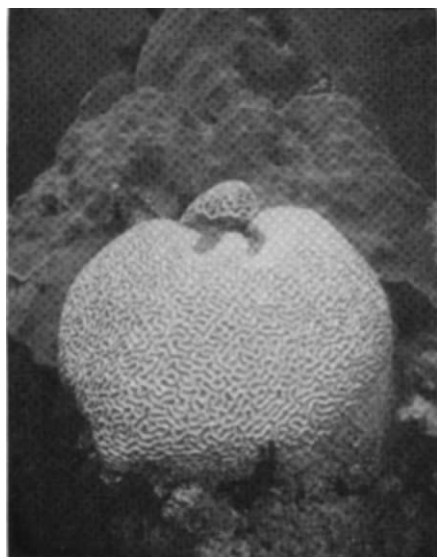
Orientation might well be reversible by structural changes brought about in the protein by a mutation. If the protein is a morphogenic pump, then in theory at least it could produce a consistent imbalance that biases the otherwise random part of the process. (Some of these ideas are summarized in Fig. 2) As far as it goes, Brown and Wolpert's suggestion is convincing, even if, at this stage of the game there is not much to get hold of.

How generally applicable might the model be? Certainly it could explain, as the authors believe it does, the production of *situs inversus* in conjoined twins of the newt *Triturus* (and in human twins), in which the viscera are always normal in one twin but often reversed in the other. On

1. Brown, N.A. & Wolpert, L. *Development* **19**, 1–9 (1990).
2. Boycott, A.E., Diver, C., Garstang, S.L. & Turner, F.M. *Phil. Trans. R. Soc. B* **219**, 239–251 (1930).
3. Davies, T.A. *Principles* **15**, 63–68 (1971).
4. Gunning, B.E.S., Hughes, J.E. & Hardman, A.R. *Planta* **143**, 121–144 (1978).
5. Layton, W.M. Jr. *J. Hered.* **67**, 336–338 (1976).
6. Annett, M. *Behav. Genet.* **8**, 227–249 (1978).
7. Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 83–93; 95–104 (1988).
8. Kauffman, S.A., Shymko, R. & Trabent, K. *Science* **199**, 259–270 (1978).
9. Richardson, J.S. *Adv. Protein Chem.* **34**, 167–339 (1981).
10. Freeman, G. & Lundelius, J.W. *Wilhelm Roux Arch. dev. Biol.* **191**, 69–83 (1982).
11. Bruckner, A., D'Eustaccio, P. & Horwich, L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5035–5038 (1989).

Coral reef bleaching alert

SIR—Goreau, in a recent Scientific Correspondence¹, drew attention to the coral reef bleaching of photosymbiotic hosts in the Caribbean in 1987–88. That bleaching event was world-wide, not just in the Caribbean, as we reported in a paper²



A totally bleached brain coral, *Colpophyllia natans*, among unbleached mushroom-shaped starcoral, *Montastrea annularis*.

published after Goreau's letter. Bleaching of organisms in coral reefs caused bleaching complexes to be formed throughout the world in 1979–80, 1982–83 and 1986–88. Each complex was made up of two or three bleaching events, with a preceding event occurring in 1979, 1982 and 1986, and major bleaching events occurring in 1980, 1983 and 1987 (ref. 2). Preceding events occur in the warm-water season toward the beginning of a short-term warming event, and major bleaching events in the normal warmwater season at the height of a short-term warming event. We suggested that preceding events could be used to predict major events.

As a result of Goreau's letter, we distributed our third bleaching summary and have so far received 40 reports of bleaching in mid-to-late 1989 from many areas of the Caribbean. This bleaching could represent a preceding event and we believe that a world-wide, major event may occur in 1990. A new El Niño Southern Oscillation (ENSO) may be beginning³. ENSOs have coincided with the past two (and most extensive) world-wide bleaching events, if this association continues, a major world-wide bleaching event could accompany the new ENSO. The timing between past major events also suggests the next may occur this year.

We believe that the world-wide coral reef bleaching cycle is caused by increased seawater temperatures produced by the combination of increased global temperatures of the 1980s, temporary warm-

ing events such as ENSOs, and seasonal warming. The progressive deterioration of inshore regions, including coral reefs, may have contributed to the intensity of the events by reducing the resilience or resistance of photosymbiotic hosts to the bleaching process.

Major marine ecological disturbances, such as bleaching, usually appear without warning and are over before adequate investigation can be performed. We hope this prediction will allow the scientific community to prepare for the event. Study of the occurrences, causes and interrelationships among events is an important challenge⁴. Predicting a world-wide marine ecological disturbance is, as far as we know, an entirely new concept.

We support Goreau's efforts to estab-

lish a Caribbean-wide research effort on bleaching through the Association of Marine Laboratories of the Caribbean. Members of the Caribbean Aquatic Animal Health Project (telephone 809 899 2048, ext. 211) are attempting to follow coral bleaching and other ecological disturbances, and would appreciate receiving information concerning these events.

ERNEST H. WILLIAMS JR

LUCY BUNKLEY-WILLIAMS

*Caribbean Aquatic Animal Health Project,
Department of Marine Sciences,
University of Puerto Rico,
PO Box 908, Lajas,
Puerto Rico 00667-0908, USA*

1. Goreau, T. J. *Nature* **343**, 417 (1990).

2. Williams, E. H. Jr & Bunkley-Williams, L. *Atoll Res. Bull.* **335**, 1–71 (1990).

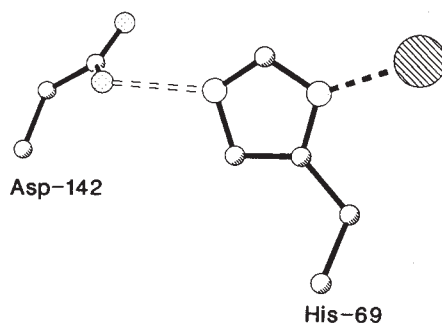
3. Monastersky, R. *Sci. News* **137**, 135 (1990).

4. Williams, E. H. Jr & Bunkley-Williams, L. *J. Aqu. Anim. Hlth* (in the press).

Another catalytic triad?

SIR—David Blow has recently commented in *News and Views*¹ on the convergent importance of the catalytic triad as observed in recently determined lipase structures. Blow, with Janet Thornton, searched the Brookhaven Protein Data Bank² for Asp[−] -- His --- Ser triads that could be related to those of the serine proteases¹. They identified two candidates (taka-amylase and immunoglobulin Kol), but each of these suffers from poor hydrogen-bond stereochemistry.

Our recent analysis of the Brookhaven



database³ complements that of Blow and Thornton. We reported that Asp[−] -- His -- Ser triads with good geometry for hydrogen bonding are present in only serine proteases. We noted that the Asp(Glu)[−] -- His hydrogen bond occurs, on average, about one per unique structure in the databank, even though some of these couples display poor geometry for hydrogen bonding. We also reported that the implicated histidine can also hydrogen-bond to a backbone carbonyl oxygen or a water molecule.

The Asp(Glu)[−] -- His couple is a component of another triad found in the zinc enzyme family. These are base-histidine couples as zinc ligands in carboxypeptidase A, thermolysin and carbonic anhydrase as noted by Liljas and Rossmann⁴. We find

that the triad Asp(Glu)[−] -- His -- Zn²⁺ is present in the three-dimensional structure of at least 36 zinc enzymes and their homologues⁵. Indirect carboxylate-zinc interactions (across bridging histidines) may contribute to biological metal-ion recognition and affinity, and such interactions could be important factors in metalloprotein folding.

We speculated¹ that the Asp[−]-His couple of the serine protease active site may be an evolutionary/structural motif related to that of the zinc protease active site. The Asp 142/His 69 couple, for example, ligands the zinc ion of carboxypeptidase A (ref. 5, see figure), and the Asp 170/His 142 couple ligands the zinc ion of thermolysin⁶. These zinc-binding motifs are intriguing in view of the fact that the Asp[−] -- His couple of the serine protease active site selectively binds transition metals, for example, as contained in platinum complexes⁷ and even the Zn²⁺ ion itself⁸. This versatile couple is certain to turn up in additional biological circumstances.

DAVID W. CHRISTIANSON

RICHARD S. ALEXANDER

*Department of Chemistry,
University of Pennsylvania,
Philadelphia,
Pennsylvania 19104-6323, USA*

1. Blow, D. *Nature* **343**, 694–695 (1990).

2. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. *J. Mol. Biol.* **112**, 535–542 (1977).

3. Christianson, D. W. & Alexander, R. S. *J. Am. Chem. Soc.* **111**, 6412–6419 (1989).

4. Liljas, A. & Rossmann, M. G. *A. Rev. Biochem.* **3**, 475–507 (1974).

5. Rees, D. C., Lewis, M. & Lipscomb, W. N. *J. molec. Biol.* **168**, 367–387 (1983).

6. Holmes, M. A. & Matthews, B. W. *J. molec. Biol.* **160**, 623–639 (1982).

7. Brothers, H. M. & Kostic, N. M. *Biochemistry* **28**, 1944 (1989).

8. Fujinaga, M. & James, M. N. G. *J. molec. Biol.* **195**, 373–396 (1987).

Speech and language defects

SIR—In a recent Scientific Correspondence¹, Gopnik presented findings on specific language impairment (as 'developmental dysphasia' is more often called) in various members of a large family. These linguistic deficits are attributed to the impairment of one particular grammatical faculty, and the problem termed 'feature blindness' by Gopnik.

'Feature blindness' is to be understood as a grammatical competence deficit, which will manifest itself across a range of linguistic performance tasks. Gopnik also suggests that the errors her subjects make could be more generally typical of language-impaired individuals. An explanatory account of specific language impairment derived from grammatical theory, and linked to genetic factors, would be an exciting development. There are reasons, however, to be cautious about Gopnik's claims.

Specific language impairment, outside the family studied, takes on more diverse forms than Gopnik envisages. Recent reports have identified, in impaired individuals within the age-range of Gopnik's subjects, problems with verb complementation², temporal adverbials³ and complex sentence formation⁴. These areas of the grammar seem to fall outside the purview of Gopnik's hypothesis, and so her explanation cannot hold for specific language impairment more generally considered.

It is not clear that the restricted type

of impairment in the affected family members can be explained only as a deficit in the underlying grammar. For example, a median score of 18 is reported for the language-impaired group (compared with 29 for normals) on a test of grammatical features (forms like plural, past tense and third-person singular). The groups do seem to be distinct on this presumably key measure of Gopnik's hypothesis.

But how is it that half of the impaired group is able to achieve a score of 60 per cent or more on this test if they are feature-blind? This is a high level of performance for an individual who has an underlying grammatical deficit in this area. It would seem more sensible to use the term 'feature-impaired', and to consider the possibility, documented elsewhere⁵, that the more likely source of the variable performance problems lies in the language-production processing system, rather than in the underlying grammar.

Grammatical forms like plural, past tense and third person singular are particularly vulnerable, in English, in individuals with phonological production prob-

lems. Was the phonological ability of the subjects reviewed, and eliminated as a factor in the impairment of some or all of them? Until alternative explanations have been explicitly considered, there must be some scepticism about Gopnik's thesis.

PAUL FLETCHER

*Department of Linguistic Science,
Faculty of Letters and Social Sciences,
University of Reading,
PO Box 218, Reading RG6 2AA, UK*

SIR—Gopnik presents¹, from a linguistic viewpoint, preliminary data on dysphasia in members of a large family. The disorder is characterized as developmental dysphasia, and the basic impairment is described as one of the underlying grammar; the impaired family members are said to be "feature blind". It is misleading to present data and conclusions based on only one aspect of the dysphasia — category-specific grammatical processing — and to neglect to report that the family has a severe congenital expressive speech disorder.

At the Institute of Child Health, a comprehensive series of investigations on the same family is in progress: genetic (M. Pembrey)², neurological (B. Neville), neuropsychological (our work) and speech (N. Jolleff, S. Jones, unpublished). The 29-member family, spanning three generations, includes 16 affected members.

Our investigations to date show that the affected members are impaired in repeating phonemes, words, non-words and sentences. Furthermore, they are impaired in aspects of language that are unrelated to syntax. For example, they make semantic errors in naming and they perform poorly on tests of receptive vocabulary.

Even within the domain of grammar, there is evidence that the impairment is pervasive and not specific to morphological markers, tenses and plural endings as suggested by Gopnik. As one illustration, on a test of receptive grammar³, the affected members show deficits in understanding many other types of syntactical structure, such as the reversible passive, postmodified subject, relative clause and embedded forms.

Gopnik has focused on only one aspect of the disorder, and it is inaccurate to conclude that the speech and language problem stems simply from an impairment in manipulating feature markers.

F. VARGHA-KHADEM

*Institute of Child Health,
Mecklenburgh Square,
London WC1N 2AP, UK*

R. E. PASSINGHAM

*Department of Experimental
Psychology,
University of Oxford,
Oxford OX1 3PS, UK*

1. Gopnik, M. *Nature* **344**, 715 (1990).
2. Hurst, J.A. *et al. Med. Child Neurol.* **32**, 347–355 (1990).
3. Bishop, D.V.M. *TROG: Test for Reception of Grammar* Department of Psychology, Univ. of Manchester, (1982).

Fractal magmatic systems

SIR—It seems reasonable to imagine a magmatic system as a complex network of fractures, veins and veinlets, ranging in size from that of the feeding dykes down to that of intergranular films. I conjecture that, like that of a blood-vessel system¹, the geometry of a magma conduit system can be described much better as a fractal² than as a euclidean shape.

Spera showed that there are thermal constraints on magma ascent^{3,4}. The rising magma gives off heat to the cool lithosphere and undergoes 'heat death' and freezing, unless the ascent velocity is high and/or the feeding dyke is large. When the melt starts to move towards the surface, the smallest veins quickly freeze and release fluids. If these are collected by the melt in an immediately larger network, the ascent of the remaining magma is easier because hydrous fluids decrease both magma density⁵ and viscosity⁶, thus tending to increase buoyancy and to decrease friction forces. The process continues all the way to the surface, and its extent depends on the local lithospheric thermal regime⁷. Heat death of a part of the magma may help the rest of the magma to make its way to the surface. The magmatic structure, like that of a root, is

expected to be more finely ramified at depth; at shallow depth it may have reduced itself to a single dyke.

The surviving magma is enriched in elements which have a strong affinity for CO₂-H₂O fluid phases. Mobile elements like radium (see refs 8, 9) are extracted at a microscopic scale from a much larger mantle volume than the major elements. Their enrichment in the final magma is a function of their affinity for the fluid phase and of the efficiency of the extraction process.

MASSIMO CORTINI

*Dipartimento di Scienze della Terra,
Università di Napoli Federico II,
Largo S. Marcellino 10,
80138 Napoli, Italy*

1. Grebogi, C., McDonald, S.W., Ott, E. & Yorke, J.A. *Phys. Lett.* **110A**, 1–4 (1985).
2. Mandelbrot, B.B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1983).
3. Bottinga, Y. & Weill, D.F. *Am. J. Sci.* **260**, 169–182 (1970).
4. Shaw, H.R. *Am. J. Sci.* **272**, 870–893 (1972).
5. Spera, F.J. *Contrib. Mineral. Petrol.* **88**, 217–232 (1984).
6. Spera, F.J. in *Mantle Metasomatism* (eds Menzies, M.A. & Hawkesworth, C.J.) 1–20 (Academic, London, 1987).
7. Bruce, P.M. & Huppert, H.E. *Nature* **342**, 665–667 (1989).
8. Rubin, H.K. & Macdonald, J.D. *Nature* **335**, 158 (1988).
9. Cortini, M. in *Uranium — Geochemistry, Mineralogy, Exploration and Resources* (eds De Vivo, B., Ippolito, F., Capaldi, G. & Simpson, P.R.) 4–11 (Institution for Mining and Metallurgy, London, 1984).

Nutrient release in leaf litter

SIR—Kundu has suggested¹ that the contribution of litter to plant nutrition at the onset of monsoon compared to microbial biomass for the dry tropical forest and savanna ecosystems is larger than we reported². We still believe that most of the nutrients needed to initiate plant growth at the onset of the rainy season in these ecosystems probably comes from mineralization of microbial cells instead of decomposing litter.

Plant growth and foliage expansion in nutrient-poor, dry tropical ecosystems occur rapidly at the start of the rainy season. Because of the lack of any conspicuous time-lag, substantial plant biomass develops within the first 2–3 weeks. This process requires a substantial flush of nutrients available for plants. It is known that in the absence of an exogenous supply of available nutrients, the net release of nutrients from most decomposing litter species occurs only after a period of net immobilization³. On the other hand, the microbial communities of droughty soils are pre-adapted to moisture stress and accumulate intracellular solutes under conditions of low water potential⁴. In soils with wet and dry seasons, increase in water potential may cause a seasonal turnover of microbial biomass by inducing microbial plasmolysis whereupon all cellular constituents, including nitrogen and phosphorus, are released⁵. Increased soil nutrients due to drying and subsequent rewetting is largely derived from dead microbial cells⁶. Expansion of microbivore populations at the start of the rainy season induces further release of microbial nutrients.

Our studies show that the pulsed turnover of microbial biomass in the early part (first 4 weeks) of the rainy season contributes about 32 kg per ha N and 13.2 kg per ha P in forest soils and about 25 kg per ha N and 10 kg per ha P in savanna soils⁷. These values compare with total rainy season (12–14 weeks) litter decomposition releases of about 22 kg per ha N and 1.4 kg per ha P for forest ecosystems and 19 kg per ha N and 1.8 kg per ha P for savanna ecosystems⁸. These release rates from litter decomposition are markedly lower compared with the values cited by Kundu from ecosystems in Zaire and Guatemala, none of which is a dry tropical site.

We suggest that the pulsed release of microbial nutrients at the beginning of the rainy season not only supports the initiation of plant growth, it also helps shorten the net nutrient immobilization phase of litter decomposition. Net nutrient mobilization during the subsequent release phase of litter decomposition sustains further biomass accumulation in the later part of the growing season.

A. S. RAGHUBANSHI

S. C. SRIVASTAVA

R. S. SINGH

J. S. SINGH

Department of Botany,
Banaras Hindu University,
Varanasi 221005,
India

Halothane and anaesthesia

SIR—Evers *et al.*¹ reported that pharmacologically relevant concentrations of halothane bind saturably to a unique environment in brain. These findings seemed to be a significant contribution to our understanding of the molecular basis of anaesthesia, an area that remains controversial^{2,3}. Recently, the same investigators published a correction¹, in which they reported that anaesthetic-induced depression of halothane uptake rather than a finite number of binding sites was responsible for the “saturable” binding of this anaesthetic in brain. Nonetheless, they suggested the short ¹⁹F NMR spin-spin relaxation time (T_2) is likely to be pharmacologically relevant. Using a similar method, we find the short T_2 environment

¹⁹F SPIN-SPIN RELAXATION TIMES (T_2)
OF HALOTHANE

Tissue	T_2 (ms)
Brain	3.9 ± 0.2 (4)
Packed red blood cells	3.0 ± 0.2 (3)
Liver	5.6 ± 0.2 (3)
Kidney	4.4; 4.1
Serum	44.0 ± 2.6 (3)

Rats were anaesthetized with 4% halothane for 1 h, decapitated, and tissues from individual animals transferred to a gas-tight syringe fitted with an 18-gauge needle. T_2 and halothane concentrations were determined by ¹⁹F NMR⁴. T_2 values represent $\bar{x} \pm$ s.e.m. with the number of experiments in parentheses. In some experiments, mixed trunk blood from anaesthetized rats was collected in NMR tubes. After centrifugation, halothane concentrations and spin-spin relaxation times of packed red blood cells and serum were determined as described elsewhere for other tissues⁴. When brain samples equilibrated with anaesthetic saturated buffer were examined without adequate centrifugation, biexponential semilogarithmic relaxation profiles (long and short T_2 relaxation times) were obtained, presumably resulting from a heterogeneity of environments. The presence of compartmental heterogeneity was also readily demonstrable in spectra of liver which exhibited two signals of comparable intensity (separated by 0.23 p.p.m.); one with a long T_2 (31 ms) and the other with a short T_2 (5.6 ms).

for halothane is not unique to brain, but is present in other tissues (see table).

We measured halothane concentrations of 2.4 ± 0.2 mM in the brain with the anaesthetic in a short T_2 environment (3.9 ms) in rats exposed to 4% (v/v) halothane for 60 min. These findings are comparable to those of Evers *et al.*¹. But we also measured short T_2 environments for halothane in packed red blood cells and other peripheral tissues of anaesthetized rats (see table).

These findings indicate that a highly immobilized environment for halothane is not unique to brain. Although the presence of this environment for halothane in brain and other tissues may be of interest, our findings indicate it is not directly related to the molecular mechanisms of anaesthesia.

HERMAN J. C. YEH

Laboratory of Analytical Chemistry,

ERIC J. MOODY

PHIL SKOLNICK

Laboratory of Neuroscience,

NIDDK, National Institutes of Health,

Bethesda,

Maryland 20892, USA

1. Evers, A.S., Berkowitz, B.A. & d'Avignon, D.A. *Nature* **303**, 157–160 (1987); correction *Nature* **341**, 766 (1989).
2. Franks, N.P. & Lieb, W.P. *Trends pharmac. Sci.* **8**, 169–174 (1987).
3. Forman, S.A. & Miller, K.W. *Trends pharmac. Sci.* **10**, 449–452 (1989).
4. Moody, E., Yeh, H.J.C. & Skolnick, P. in *Alcohol Reinforcement* (eds Koob, G. *et al.*) (Birkhauser, Boston, in the press).

■ An earlier version of this letter was received on 14 December 1988.

Beach pebbles explained

SIR—Wald (*Nature* **345**, 211; 1990) seems unaware that the presence of a planar structure in a rock will strongly influence the final proportions of a beach pebble. Most rocks contain structures such as sedimentary bedding, metamorphic foliation or igneous lamination, and will be shaped to flattened ovoids of the type he describes, the plane of flattening being parallel to the planar structure.

This is easily verified by direct observation on almost any beach. Even in apparently structureless rocks such as quartzite there is likely to be a preferred orientation of the crystallographic axes of the quartz crystals such as to cause a biased pebble shape. To avoid such complications, one has to be choosy about the rock: a fine-grained intrusive igneous rock might do, such as the microgranite of Ailsa Craig in Scotland, long used in the manufacture of curling stones because of its small grain size and homogeneity.

W. ASHCROFT

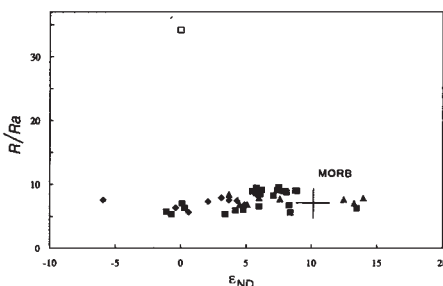
Department of Geology and
Petroleum Geology,
University of Aberdeen,
Marischal College,
Aberdeen AB9 1AS, UK

1. Kundu, D.K. *Nature* **344**, 203 (1990).
2. Singh, J.S., Raghubanshi, A.S., Singh, R.S. & Srivastava, S.C. *Nature* **338**, 499–500 (1989).
3. Upadhyay, V.P. & Singh, J.S. *J. Ecol.* **77**, 127–146 (1989).
4. Sparling, G.P., Milne, J.G. & Vincent, K.W. *N.Z.T. Agric. Res.* **30**, 79–84 (1987).
5. Kieft, R.L., Soroker, E. & Firestone, M.K. *Soil Biol. Biochem.* **19**, 119–126 (1987).
6. Sparling, G.P. & Ross, D.J. *Plant Soil* **105**, 163–167 (1988).
7. Srivastava, S.C. thesis (Banaras Hindu Univ., Varansi, 1989).
8. Singh, K.P. & Misra, R. *Structure and Functioning of Natural, Modified and Silvicultural Ecosystems of Eastern Uttar Pradesh*. (Tech. Rep., Banaras Hindu Univ., 1979).

Volcanic traces

SIR—Numerous chemical and isotope studies^{1,2} of xenolith samples carried to the surface by volcanic eruptions have revealed a common pattern of enrichment in incompatible trace elements, accompanied by only minor amounts of basaltic components such as silicon, aluminium, calcium and titanium. Termed 'metasomatic' enrichment (especially where the addition of the light rare-earth elements is involved), this process has been attributed to the action of mantle fluids enriched in trace elements, capable of modifying the content of trace elements in rocks through which they pass without discernible effect on the composition of major elements². Support for this view comes from the common occurrence of CO₂-rich fluid inclusions in xenoliths and experimental indications of high rare earth solubility in both carbonate melt and CO₂ (refs 3,4). However, measurements of trace element solubilities in CO₂ at high pressure and temperature remain in conflict. Meen *et al.* recently⁵ failed to substantiate the earlier experimental results: their experiments suggest distribution coefficients $K_D^{CO_2/diopside}$ for Nd at least as low as 0.5 and probably less than 0.05 at high pressures and temperatures, implying a limited role for CO₂-rich fluids as agents of rare-earth element transport in the mantle. Our isotope measurements on xenolith samples support the experimental conclusions reached by Meen *et al.*⁵.

Helium isotopes are ideal tracers of fluid movement in the crust and mantle, partitioning strongly into fluid in equilibrium with melt or solid phases⁶. Helium in xenolith samples resides predominantly in the (variably CO₂-rich) fluid inclusions⁷⁻⁹. The isotope covariation of neodymium with He provides a test of the fluid-phase partitioning behaviour of Nd — if mantle fractionations of U/He and



He—Nd isotope relationships in ultramafic xenoliths. The wide range of Nd isotope compositions (expressed as $\epsilon_{Nd} = 10^4 \left(\frac{^{143}Nd}{^{144}Nd} \right)_{sample} - 1$, where CHUR indicates a reservoir with chondritic $^{147}Sm/^{144}Nd$ ratio) contrasts with uniform, mid-ocean-ridge-basalt-like He isotope compositions (sample $^3He/^4He$ ratios (R) normalized to the atmospheric ratio (R_a)). Note that radiogenic He plots at low R/R_a values (less than 0.1). Diamonds, continental samples; triangles, convergent margin samples; closed squares, ocean island samples; open square, primitive components.

Sm/Nd remain coupled, and Nd were to move with He in metasomatic fluids, He and Nd isotope signatures should vary sympathetically.

He and Nd isotope data for ultramafic xenoliths have been reported by us⁷⁻⁹: in the figure we illustrate these results for an extensive set of continental and oceanic samples. Over an extreme range of sample Nd isotope compositions, He maintains a strikingly uniform composition indistinguishable from that of the convective mantle as sampled by mid-ocean ridge basalts. The wide-ranging Nd isotope compositions of the xenoliths may result from inputs from other geochemical reservoirs⁸, or from long-term isolation in the lithosphere. However, they are decoupled from He isotope effects. Neither primitive 3He -rich, nor 4He -rich compositions indicative of ageing in the lithosphere, are observed in the samples. The He and Nd carried by the samples come from separate reservoirs, with separate fractionation histories. The data imply that these elements do not move together in metasomatic fluids.

Decoupling of these systems may be thought of as reflecting addition to the lithospheric mantle of a He-rich, Nd-poor component. The high He component is likely to have been associated with CO₂,

now present in fluid inclusions.

The xenolith data document widespread access of volatiles from the convecting mantle to the lithosphere beneath regions of alkaline volcanism. They support the results of Meen *et al.*⁵, suggesting that these CO₂-rich fluids are not significant carriers for the rare earth and perhaps other trace elements, such as strontium⁷⁻⁹, in the mantle lithosphere.

J. STONE
D. PORCELLI
D. VANCE
S. GALER
R. K. O'NIONS

Department of Earth Sciences,
University of Cambridge,
Downing Site,
Cambridge CB2 3EQ, UK

1. Frey, F.A. & Green, D.H. *Geochim. cosmochim. Acta* **38**, 1023–1059 (1974).
2. Menzies, M.A., Rogers, N., Tindle, A. & Hawkesworth, C.J. in *Mantle Metasomatism* (eds Menzies, M.A. & Hawkesworth, C.J.) 313–361 (Academic, London, 1987).
3. Mysen, B.O. *Neues Jb. Miner. Abh.* **146**, 41–65 (1979).
4. Wendlandt, R.F. & Harrison, W.J. *Contrib. Mineral. Petrol.* **69**, 409–419 (1979).
5. Meen, J.K., Eggler, D.H. & Ayers, J.C. *Nature* **340**, 301–303 (1989).
6. Jambon, A. *Chem. Geol.* **62**, 131–136 (1987).
7. Porcelli, D.R., O'Nions, R.K. & O'Reilly, S.Y. *Chem. Geol.* **54**, 237–249 (1986).
8. Vance, D., Stone, J.O.H. & O'Nions, R.K. *Earth planet. Sci. Lett.* **96**, 147–160 (1989).
9. Porcelli, D.R., O'Nions, R.K., Galer, S.J.G., Cohen, A.S. & Matthey, D.P. *Contrib. Mineral. Petrol.* (in the press).

Hepatocytes and scatter factor

SIR—Scatter factor, a cytokine secreted by certain fibroblasts, enhances the movement and causes the dissociation and scattering of epithelial cells¹⁻³, and may be involved in epithelial migration, for

cell motility.

If the scattering and mitogenic activities are due to the same protein, it is tempting to suggest that different classes of receptors mediate the effects on growth and

Mouse scatter factor

Rat hepatocyte growth factor
Rabbit hepatopoietin A
Human hepatocyte growth factor
Human hepatocyte growth factor

VVNGIPTQTTVG?MVSLLYRN?HI

-----W-----K---K---
---K---RT-V-R---K---K---
-----R-NI-W---R---K---
-----R-NV-W-I---R---K---

Ref.

4
5
6
4

example in embryogenesis². Mouse scatter factor has an apparent relative molecular mass of 62,000 (62K) in non-reducing SDS gels and is composed of two subunits of 57K and 30K held together by disulphide bonds³. We have obtained an amino-terminal sequence of the 30K subunit and notice that it is very similar to that of the smaller subunit of hepatocyte growth factor/hepatopoietin A, a potent mitogen for rat hepatocytes which has been implicated in liver regeneration *in vivo*.

The sequence data suggest that scatter factor and hepatocyte growth factor/hepatopoietin A are either the same or highly related molecular species. Either way, the relationship is intriguing. Scatter factor was identified as a motility factor and, at least in MDCK cells which are highly responsive to motility stimulation, it has little effect on DNA synthesis and no effect on the rate of cell division². On the other hand, hepatocyte growth factor/hepatopoietin A has no reported effect on

movement in different types of cells or tissues. If they are due to related but different proteins, it is clear that the two factors are members of a new group of cytokines that control both epithelial movement and growth. Coordinate regulation of cell migration and division is a key feature of many developmental processes and repair reactions.

ERMANN GHERARDI
MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK
MICHAEL STOKER
Imperial Cancer Research Fund and
Department of Pathology,
University of Cambridge,
Cambridge CB2 1QP, UK

1. Stoker, M. & Perryman, M. J. *Cell Sci.* **77**, 209 (1985).
2. Stoker, M. *et al. Nature* **327**, 239–242 (1987).
3. Gherardi, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 5844–5848 (1989).
4. Nakamura, T. *et al. Nature* **342**, 440–443 (1989).
5. Zarnegar, R. *et al. Biochem. biophys. Res. Commun.* **163**, 1370–1376 (1989).
6. Miyazawa, R. *et al. biochem. Biophys. Res. Commun.* **168**, 967–978 (1989).

A habit of collaboration

Philip Gummett

Alvey. Britain's Strategic Computing Initiative. By Brian Oakley and Kenneth Owen. MIT: 1990. Pp.337. £24.95, \$40.

THREE quotations encapsulate the key question raised by this book: "... what's this bloody nonsense? We don't have strategies in this government." "I wish to argue for the development in a very short space of time of a national strategy in which the government has to take the lead.... The opportunity for Britain in this industry is immense, and we must not let it slip between our fingers." "This lack of any long-term view worries me. It's the difference between running a corner shop and running a major chain of stores."

The first was Lord Trenchard, industry minister under Sir Keith Joseph at the UK Department of Industry in 1979, in response to a departmental report entitled "A strategy for information technology". The second was Kenneth Baker in 1980, then a backbench member of parliament, in a speech in which he proposed a ten-point programme for information technology in Britain. One recommendation was for the establishment of a minister of information technology — a post that was created, and to which Baker himself was appointed in 1981. The third quotation was David Talbot, former Alvey software engineering director, commenting on the decision not to mount a specific follow-on programme to Alvey.

For a nation of corner shopkeepers, in the ideological climate of the 1980s, the real mystery is how the national strategy for information technology that was embodied in the Alvey programme ever got off the ground. It was unsurprising that proposals for a follow-on, prepared in 1986 by the Bide committee, should have lain inside Whitehall for 13 months before being dismissed in a few lines in a white paper (policy document), the main purpose of which was to describe the transformation of the Department of Trade and Industry into the "Department for Enterprise".

The Alvey programme was a unique experiment in technology policy in Britain. It involved expenditure of £350

million over the financial years 1984–89 in the fields of very large-scale integration (VLSI), software engineering, intelligent knowledge-based systems (artificial intelligence), and man-machine interfaces, together with some large-scale demon-

committed, the programme comprised more than 200 cooperative projects, each typically having two or three companies working with one or two universities. Participants included 115 firms, almost every university, 19 polytechnics and about 20 government research establishments and research associations. At the peak, around 2,500 people were involved, most from industry. By 1987, 10 out of about 200 industrial projects had put products on the market or had improved existing production processes, and a further 77 projects had products at the prototype stage.

Whatever technologies may emerge from Alvey, its main legacy will lie in the vitality of an expanded information technology community that spans the industry-government-academia triangle, and whose members are capable of, and used to, working with each other. The habit of collaboration has even led the Department of Trade and Industry and the Science and Engineering Research Council to integrate their committee structures for post-Alvey information technology policy, and to express the hope that the Ministry of Defence might join them. As a springboard for entry into European projects, particularly ESPRIT, where the centre of gravity of UK support for information technology research now lies, Alvey has served the national community well. Whether a second national programme, running alongside ESPRIT, as recommended by Bide, would have served the country even better, will remain a matter for debate.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Brian Oakley — compelled to "expose his activities to scrutiny".

strator projects. Of the total, £200 million came from three government sources (the Ministry of Defence, which was particularly interested in the VLSI work, the Department of Industry and the Science and Engineering Research Council), and a further £150 million from industry. But collaboration was not confined to sources of financial support: rather, it was the bedrock of the entire initiative. For a project to be supported, it had to involve collaboration between firms, higher-education establishments or government research establishments. By pooling resources for the funding and performance of pre-competitive research, the aim was to propel Britain to the point where it could compete on the world stage in fifth-generation computing.

By the time all the funds were

Running such an enterprise takes unusual skills. The director selected was Brian Oakley, then secretary of the Science and Engineering Research Council. His earlier experience included computing at the Royal Signals and Radar Establishment and at the headquarters of the Ministry of Technology, followed by setting up the research requirements boards at the Department of Industry after the Rothschild reforms of the early 1970s. So he knew the ways of scientific administrators and government departments, and was used to drawing industrialists into setting government research strategies.

Another feature of Oakley's make-up distinguishes him from the stereotypical civil servant. This is his compulsion to expose his activities to scrutiny and criticism, including self-criticism. His

co-authorship of this book, together with Kenneth Owen — journalist and editorial consultant to the Alvey directorate — is entirely in character in its detailed frankness about the problems and failings as well as the achievements of the Alvey programme. So also was his decision, which was novel at the time in Whitehall, to employ independent evaluators, charged with feeding back their assessments periodically throughout the programme.

The book is highly informative on the manoeuvrings within the information technology community that brought an ideologically ill-disposed government to the point of contributing £200 million of taxpayers' money to the programme. Analysts interested in the social construction of technology or in the operation of policy networks will find much of value here, as will readers with more general interests in how governments can be moved. The same richness of detail, sense of excitement and bluntness in assess-

ments characterizes the discussions of how the programme was structured and administered, its development and achievements, and the failure of the Bide proposals.

If I have a criticism, it is that the use of quotations to capture the recollections and judgements of participants is overdone. It leads to repetition and, at times, unclear focus. The journalist's bias towards the provocative quotation sometimes wins out over the reader's desire to know whether, for example, a citation was typical or not of the people consulted. A tougher final editing would have been worthwhile. But this book does not claim to be the last word on Alvey. Its aim — to provide a personal memoir and to convey the flavour of the programme — is amply fulfilled, and later writers on Alvey will find it an invaluable source. □

Philip Gummett is in the Department of Government, University of Manchester, Oxford Road, Manchester M13 9PL, UK.

An atmosphere of thought

Priscilla C. Grew

Science and Russian Culture in an Age of Revolutions. By Kendall E. Bailes. *Indiana University Press: 1990. Pp. 238. \$29.50.*

THE authors of a book on global change, *History of the Earth's Atmosphere* (Springer-Verlag, 1987), concluded: "controlling the chemical composition of the atmosphere . . . will be a great step towards the formation of the noosphere; and this will in turn realize the idea of the founder of biosphere science, V. I. Vernadsky". The authors, M. I. Budyko, A. B. Ronov and A. L. Yanshin, did not find it necessary to explain what "noosphere" means, nor who V. I. Vernadsky is. Many readers west of Berlin, however, will experience some hesitation in defining the word "noosphere", and few, unless they frequent the New Age bookstores, will recall ever having read anything written by the Ukrainian geochemist V. I. Vernadsky (1863–1945). At present there is only one paperback by Vernadsky in print in the United States, a translation of *The Biosphere* (written in 1926), published by the Synergetic Press in Oracle, Arizona.

The year 1988 was the 125th anniversary of Vernadsky's birth. For the occasion, the vice-president of the USSR Academy of Sciences wrote an article stating that Vernadsky "stands on a par" with Aristotle and Avicenna. Scientific conferences in Vernadsky's honour were held in Moscow, Leningrad, Kiev

and East Berlin. Also in 1988, Kendall Bailes, who died of AIDS at the age of 48, finished this first major biography of Vernadsky in English.

Many in the West have dismissed Soviet adulation of Vernadsky as icon-making by a society with a penchant for selectively quoting higher authorities "from the tomb" to justify political ends. Vernadsky was indeed a victim of the stalinist policy of "perish then publish". Many of his works were not printed until decades after 1945, and even then they appeared in a confusing array of variously censored editions. Vernadsky's copious letters, diaries and private papers have long been inaccessible to Western scholars. For almost 50 years, an official Soviet commission to study the scientific heritage of V. I. Vernadsky has been hard at work.

Bailes, author of the prizewinning *Technology and Society under Lenin and Stalin*, was fully competent to cut through these problems. His concise book, with references that stop short of the Gorbachev era, will be the foundation for all future scholarship in English on Vernadsky. Meticulously referenced with a wealth of primary sources, the book clearly documents Vernadsky's scientific achievements, such as his founding of the discipline of biogeochemistry and his promotion of biosphere science. Both these fields are increasingly critical for understanding global change.

Bailes also puts into context the noosphere business. Vernadsky, himself an historian of science, was careful to give credit to the mathematician Edouard Le Roy for originating the term "noosphere". Vernadsky's later development of the concept diverged from Teilhard de Chardin's mystical noosphere of a collective

global consciousness, like an "envelope of thought". Vernadsky foresaw that inadvertent human interference in global biogeochemistry would gradually be succeeded by intentional human shaping of global change. He called this transition the evolution from the biosphere to the noosphere. For Vernadsky, the noosphere emerges as humans assume an ever greater role in managing, rather than "unconsciously" inducing global change. In Vernadsky parlance, our plan to diminish the hole in the ozone layer is a clear symptom of the emergence of the noosphere.

For Bailes, the central jewel of Vernadsky's legacy was his life-long "assertion of the primary value of freedom of thought for science and for human creativity" (p. 197). The archival record reveals that Vernadsky steadfastly sustained this assertion through revolution and war, arrest, censorship, and the disappearance and murder of students and friends. Bailes shows that Vernadsky advocated representative government and federalism, abhorred violence and courageously defended colleagues from political repression.

Does Vernadsky rank with Aristotle? Probably not, but perhaps the parallel to Avicenna (ibn Sina of Bukhara) is not so far-fetched. Both had catholic interests in science and philosophy, and it did take two centuries for ibn Sina to get his name and work translated into the Latin world of the West. A posthumous paper published in *American Scientist* in 1945, "The Biosphere and the Noosphere", is Vernadsky's only journal article in English. His striking portrait is the frontispiece. Like a chiaroscuro "St Jerome in the Study", Vernadsky, with long silver hair, white beard and downcast eyes, rests his bowed head on his clenched fist. Beneath the photograph are a few typewritten lines in Russian, cut from a letter to a friend. "I look forward very optimistically", he writes. He doesn't appear optimistic in the picture — but icons never smile. □

Priscilla C. Grew is at the Minnesota Geological Survey, University of Minnesota, 2642 University Avenue, St Paul, Minnesota 55413, USA.

Environmental paperbacks

■ Now available in Bantam paperback is James Lovelock's *The Ages of Gaia: A Biography of our Living Earth*. For review see *Nature* **336**, 270; 1988. Price is \$10.95.

■ The third edition of *The Human Impact on the Natural Environment*, by Andrew Goudie, was recently issued by Blackwell. Price is £10.95. (A hardback edition is available at £30.)

■ Published on 5 July by Penguin, *A Green Manifesto for the 1990s* by Penny Kemp and Derek Wall explains Green party policy in Britain — not only its environmental programme but also medical, defence and social issues. Price is £4.99. □

Formed for function

H. C. Bennet-Clark

Animals. By R. McNeill Alexander. Cambridge University Press: 1990. Pp.509. Hbk £50, \$89.50; pbk £19.50, \$34.50.

THE ways in which animals go about their business of catching food, moving about, protecting themselves, reproducing, growing, getting rid of unwanted material and the general housekeeping of their lives offer a fascinating series of problems for study by zoologists. There is, among animals, a bewildering variety of body plans, detailed differences in design between closely related groups and, in terms of diversity of species or abundance, great differences in the apparent success of different groups.

The classical approach to the study of animals has long been served by textbooks that offer a systematic framework, describe the comparative anatomy and, group by group, consider aspects of the biology of various parts of the animal kingdom. The distinction between vertebrates and invertebrates that appears at the level of the student textbook is traditional and, like many traditions, has an emotional basis: we are vertebrates — and mammals to boot — so we tend to feel more empathy with, and to be more intrigued by, rats or cats than by frogs or fishes, and to find squids, snails or jellyfish positively alien.

This ranking may blinker us from noting that such distinct groups as insects or mammals face similar problems; of desiccation or of living in a medium that does not support their bodies. Such problems share physical or chemical bases so it is hardly surprising that the two otherwise distant groups solve similar problems with somewhat similar mechanisms, exemplified by the evolution of water-retention mechanisms or of jointed legs in both arthropods and tetrapod vertebrates. The pathways by which they have evolved are very different but the ways in which legs are used are similar — crabs trot as do horses, albeit sideways and sometimes with eight or six legs.

Students of the comparative approach to animal form and function have been well served by two earlier textbooks by Alexander, *The Chordates* and *The Invertebrates*, which provide excellent accounts of how animals work, always with a systematic framework but with far less attention to details of classification or comparative anatomy than appear in other major textbooks. In my view, these two books offered the clearest most accessible and modern accounts of how animals live and function. The weakness

of the approach is that it doesn't tell you much about what the animals are — but contrariwise, the classical approach tends to offer frustratingly little about how the animals work.

Animals, according to Alexander's own preface, covers the same ground as *The Invertebrates* and *The Chordates* in roughly half the number of pages. It treats all animal-like organisms from protists to mammals. It is aimed at elementary undergraduate courses, as an adjunct and complement to the more classical textbooks but also as a central textbook.

How successful is the new approach? I have always admired the clarity of

emphasis on vertebrates than on invertebrates and may here be justified because the phylogenetic links and succession in vertebrates is well understood whereas that of invertebrates is a mess.

Alexander's great success is the way in which he draws parallels across the animal kingdom. Problems of respiratory exchange in water are compared between *Octopus* and fishes, of gaseous exchange between insect flight muscles and bird lungs, and so on. This approach is excellent because it bridges the traditional but artificial and dangerous gap that is fostered by the clumping of animals into invertebrates and vertebrates. The treat-

ment is often uncompromising — you may have to work through calculations to follow a fine series of explanations and insights into how animals work.

Refreshing, too, is the amount of new material. Some old favourites, such as why flatworms have to be thin and flat to allow adequate gas exchange persist, but discussion of reptile feeding mechanisms or of the probable posture and speed of dinosaurs appear for the first time. Other topics receive a major update: the explanation of the function of fish swim bladders has been radically revised and improved, as have the accounts of bird and insect flight. There are also useful treatments of basic features of the embryology of worms, echinoderms and of various groups of vertebrates, serving, as always, Alexander's purpose of showing how these processes work and rarely stepping onto the minefield of drawing phylogenetic inferences. His main concern seems to be with the present problems of life and the solutions shared between animal groups or evolved uniquely as a key feature of a particular group: only where it is

germane to how the animals work does he consider the history and evolutionary stages, as with the evolution of protrusible jaws of bony fish, or of jaw mechanisms from fish to amphibians, reptiles and mammals.

As with Alexander's earlier books, this is a *tour de force*, showing great understanding and with lucid explanations. I shall be recommending it warmly to my students — but I shall also remind them of the existence of its predecessors so that they can catch and benefit from the author's many earlier insights, inevitably omitted from this briefer and re-shaped account. □

H. C. Bennet-Clark is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

New light on bird flight.

Alexander's earlier textbooks, cited above, and have recommended them warmly, but I have warned students to look elsewhere for details of what the animals are or of their anatomy. *Animals* provides even less about the anatomy or interrelations though, in fairness, it does give extensive references from which this information may be obtained. This, too, is not about what animals are but about how they work.

It is an irritating feature of the treatment that certain animal groups, like the sponges, nemertines or Onychophora, at the borders between major groups of animals, are dismissed in a few words or paragraphs, but lungfishes and *Archaeopteryx* get far more detailed treatment. Partly, this reflects Alexander's greater

Taking a close look at death

Roy Porter

Expectations of Life. A Study of the Demography, Statistics, and History of World Mortality. By H. O. Lancaster. Springer-Verlag: 1990. Pp. 605. £114, \$159.

MEDICAL scientists and historians both have a stake in understanding the role of disease in the rise of civilization, but they often approach the problem with very different preconceptions. The medical mind anticipates encountering solid evidence (archival, skeletal) upon which to train its technical skills (clinical, epidemiological, palaeopathological and so forth), and is perhaps disposed to view the springs of progress rather biologically. By contrast, the historian is probably in the first place less optimistic about our capacity even to identify the pestilences of the past, and in any case will be anxious to stress multifactoriality, the role of culture as well as nature. Politics, morals, and wealth distribution often have far greater impact than the transmigrations of microorganisms upon morbidity and mortality.

Each view has its point, and not least of the virtues of H. O. Lancaster's magisterial survey of world mortality lies in its sensitivity to both. As may be expected from an MD who is an emeritus professor of mathematical statistics, epidemiology constitutes the core of this project. Professor Lancaster devotes more than half his book to a meticulous discussion of the spectrum of disease types (bacterial, virus, arthropod-borne, neoplastic and so on) which have primarily determined death rates. But, for all his array of graphs and statistics, he never loses sight of the drastic deficiency of clinical information and mortality statistics for three-quarters of the globe until the past century. And he shows due caution regarding our capacity to establish fine retrospective diagnostic distinctions, especially amongst the numerous disenteric infections that decimated the great conurbations before the advent of compulsory public health and scientific therapeutics: after all, as he eruditely demonstrates, the stabilization of disease terminology is itself a product of modernity. (Exceptionally to his wise scepticism, Lancaster is confident that the 'great plague' of Athens can be put down to smallpox; most historians of ancient diseases would demur.)

After surveying the great diseases, Lancaster examines the other main causes of mortality (famine, war and so on) in the light of available demographic statistics, before posing the key question: how does the history of diseases influence social change, and thus shape the patterns of world population? Ever since Malthus, these variables have been fiercely disputed. Malthus pictured biology (disease,

famine) checking the progress of society. Many modern demographers have stood Malthus on his head, arguing that society itself has been the cradle of disease (through malnutrition no less than insanitary environments). Nowadays, though 'agriscience' may have sprung the malthusian trap, it is surely creating new ones in its place, quite unknown to Malthus — above all, global environmental degradation. How does Lancaster assess these variables?

He contends — surely correctly — that the history of the past 500 years reveals sharply divergent patterns for the developed and the undeveloped worlds. In the West, commercialization, industrialization and urbanization first exacerbated epidemic mortality, but subsequently provided the abundance and the technical know-how to eliminate epidemics, while artificially containing population growth. By contrast, the tale of the developing countries has been one of diminishing real

mastery over the environment (sometimes masquerading as 'agricultural improvement'), and the breakdown of traditional demographical equilibria. The successes of tropical medicine have created fresh problems in their wake (witness AIDS in central Africa).

Lancaster is sometimes content to let his material speak for itself, and a more energetic evaluation of rival interpretations would have been welcome: for instance, the path-breaking revisionist reading by Ester Boserup of *Population and Technological Change* (University of Chicago Press, 1981) is surprisingly absent from his otherwise spectacular 100-page bibliography. This points to a certain tension in the conception of this book — is it intended mainly as a work of synthesis or of reference? Sometimes the itch to be exhaustive gets the upper hand, as when a paragraph on bridge disasters is followed by a computation of mortality from lightning. Nevertheless, not least thanks to Lancaster's unfailing ability to convey highly technical matters in a lucid prose, this is a volume which is sure to be widely consulted by all concerned with the history of mortality. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BP, UK.



The first, and best known, of the Cottingley fairy photographs. In 1917 Elsie Wright (aged 15) and Frances Griffiths (aged 9) constructed the photographs using cardboard cutouts held in place with hat pins — in this photograph the girls forgot to attach wings to the fairy second from the left. The photographs continued to impress until 1983 when Wright and Griffiths revealed the manner of fabrication. Reproduced from *Fake? The Art of Deception* edited by M. Jones, the photographs are part of a collection of fakes from all fields portrayed in the book; published to coincide with the exhibition at the British Museum (9 March until 2 September). The volume contains 'evidence' of a merman, furry trout and unicorn's horn but also includes chapters on scientific detection of fakes and the limits of expertise. Published by British Museum Publications, price hbk £25; pbk £16.95. Published by the University of California Press in the United States, \$49.95; \$24.95. □

A cosmic book of phenomena

P. J. E. Peebles & Joseph Silk

A comparison of the merits of five general theories for the origin of galaxies and large-scale structure in the Universe with 38 observational constraints from extragalactic astronomy produces no clear winner. Two theories, cold dark matter in an inflationary cosmology and baryonic dark matter in a low-density Universe, emerge slightly ahead of the pack.

THERE is a tendency at scientific meetings, when a particularly important but tentative result is presented, to demand of one's colleagues what odds they would give for eventual confirmation. Many bottles of the finest champagnes and malt whiskies, and even more esoteric stakes, rest in abeyance while observers struggle to count rare photons from remote galaxies or experimentalists devote decades to designing new types of detectors. To enrich, enlighten and even amuse those of our colleagues who are trying to assess the merits of the rival cosmogonies, we have begun a modest programme of setting up a cosmic book of odds. Our first book¹ focused almost exclusively on the large-scale structure of the Universe. This one is devoted to the observable phenomena that theorists customarily invoke (or ignore) in developing models for the formation of the galaxies. Which observations are reliable? Which observations can be used to discriminate between alternative theories? As issues in extragalactic astronomy rarely are settled by any one measurement, an array of assessments, such as that in Table 1, may be what finally leads us to a true standard model for the origin of galaxies and the large-scale structure of the Universe.

We remind punters that the redshift z of an object is defined so that $1+z$ is the ratio of the observed wavelengths of spectral features to the corresponding wavelengths in the rest frame of the object. At $z \gg 1$, the redshift is a convenient measure of the epoch at which the observed radiation left the object. At $z \ll 1$, redshift is proportional to distance. Hubble's constant, the ratio of redshift to distance, is still quite uncertain, so it is written as $H = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$, where the parameter h is thought to be in the range 0.5–1. The value of h is very important for some cosmological tests, but at the crude level of our table it is only a scaling factor.

Theories

Of critical importance in the table of odds is the cosmological density parameter Ω , defined as the ratio of the mean mass density in the Universe to the value in the scale-invariant Einstein-de Sitter universe, $\Omega = 3H^2/8\pi G$, where G is the gravitational constant. Luminous matter (both stars and gas) contributes $\Omega \approx 0.007$. Mass estimates based on the dynamics of systems of galaxies usually indicate $\Omega \approx 0.1$ (ref. 2). Theorists almost all agree that Ω is probably close to unity. First, inflation makes the spatial curvature flat to within 0.01%, which implies $\Omega = 1$ unless there is a cosmological constant. For those unconvinced by the case for an inflationary Universe, there is a fine-tuning argument: the Friedmann-Robertson-Walker equation implies that the ratio of the mean mass density to the critical Einstein-de Sitter value (which is proportional to the curvature term and is at present of the order of unity) was equal to unity to within 1 part in $\sim 10^{60}$ at the Planck epoch, which seems unreasonable. If one is willing to live with a cosmological constant, one has a similarly spectacular coincidence for its value at the end of inflation. Moreover, there are two arguments against the idea that the 'missing mass' needed to make $\Omega = 1$ is ordinary (baryonic) matter: it is difficult to understand how

so much material could be hidden from observation, and a high baryon density would make it difficult to reconcile the calculated abundances of the light elements produced in the hot Big Bang with observed abundances³. On the other hand, there is rough agreement between the baryon density needed to account for light-element nucleosynthesis with the mass density required to account for the dynamics of galaxies. The two popular interpretations are that the mass of the Universe is dominated by some sort of exotic non-baryonic weakly interacting matter, or that $\Omega \approx 0.1$, theoretical arguments notwithstanding, and the mass density predominantly baryonic.

In the cold dark matter (CDM) theory^{4,5}, an Einstein-de Sitter universe is dominated by weakly interacting matter with negligible primeval pressure (that is, cold matter), and structure grows by gravitational instability out of a scale-invariant spectrum of primeval adiabatic density fluctuations (where the primeval ratios of local number densities of photons, baryons and dark matter particles are independent of position). This theory requires biasing, in the sense that mass is less strongly clustered than are galaxies (so that the relatively small masses per galaxy found in studies of the dynamics of systems of galaxies can be consistent with the high net mass per galaxy implied by $\Omega = 1$). The degree of biasing required depends to some extent on the phenomenon. For the sake of the argument, we assume that the r.m.s. fluctuation $\delta N/N$ in the galaxy count in a randomly placed sphere of radius equal to the galaxy clustering length $r_0 = 5.4(\pm 1)h^{-1} \text{ Mpc}$ is twice the r.m.s. fluctuation $\delta m/m$ in the mass found in the sphere.

In the hot dark matter (HDM) model⁶, the particles of dark matter have a primeval velocity characteristic of a neutrino of mass $\sim 30 \text{ eV}$. (Using this mass, and the predicted mean number density of neutrinos left over from the hot Big Bang, the mass density in neutrinos makes $\Omega \approx 1$.) Structure grows by gravity out of adiabatic fluctuations in the mass distribution. In the linear-perturbation approximation smoothing by the free streaming of the dark matter produces a mass coherence length of $\sim 5h^{-2} \text{ Mpc}$ at the present epoch.

There are two classes of models in which structure formation is seeded by primeval nonlinear perturbations. In the string (STR) theories, density perturbations are produced by the gravitational effect of a network of cosmic strings that untangles as the Universe expands, so that at any epoch t the Universe within the Hubble length ct contains a few long strings and closed loops. Recent calculations have lowered the expected abundance of massive loops relative to long strings⁷. Our impression is that the implied modification in the picture for galaxy formation, from accretion onto isolated loops to fragmentation of the wakes of long strings, makes the string-seeded galaxy formation models more attractive.

In the explosion (XPL) picture⁸ locally inserted energy, perhaps from early supernovae, piles baryons into ridges which collapse to produce new star clusters. We know that non-gravitational processes such as explosions and cooling have an essential role in the behaviour of the interstellar medium and in star

TABLE 1 Critical extragalactic phenomena, real and chimerical

	Weight	Cold dark matter		Hot dark matter		String perturbations		Explosion perturbation		Baryonic dark matter	
		p	r	p	r	p	r	p	r	p	r
1. Cosmology											
a, $\Omega \approx 0.1$	0.7	0.05	0.18	0.05	0.18	0.3	0.36	0.5	0.50	0.95	0.82
b, $\Omega_{lum} \approx 0.1 \Omega_{dyn}$	1.0	0.1	0.10	0.1	0.10	0.1	0.10	0.7	0.70	0.95	0.95
2. Cosmic background radiation											
a, Isotropy: $\delta T/T < 1 \times 10^{-4}$ at ≤ 10 arcsec	1.0	0.95	0.95	0.5	0.50	0.3	0.30	0.1	0.10	0.2	0.20
b, Isotropy: $\delta T/T < 2 \times 10^{-5}$ at ~ 30 arcmin	1.0	0.95	0.95	0.05	0.05	0.7	0.70	0.7	0.70	0.7	0.70
c, Spectrum: Compton $y < 0.001$	1.0	0.95	0.95	0.5	0.50	0.4	0.40	0.1	0.10	0.95	0.95
3. High-redshift phenomena											
a, There are quasars at $z \approx 5$	1.0	0.1	0.10	0.05	0.05	0.95	0.95	0.95	0.95	0.95	0.95
b, Intergalactic H I clouds are weakly clustered at $z \approx 2.5$	1.0	0.95	0.95	0.05	0.05	0.5	0.50	0.5	0.50	0.05	0.05
c, The quasar abundance peaks at $z \approx 2.5$	1.0	0.5	0.50	0.1	0.10	0.1	0.10	0.1	0.10	0.05	0.05
d, The metagalactic ionizing-radiation field is present at $z \approx 5$	1.0	0.5	0.50	0.05	0.05	0.95	0.95	0.95	0.95	0.95	0.95
e, There are young (~ 0.3 Gyr) galaxies at $z \approx 3$	0.6	0.95	0.77	0.8	0.68	0.5	0.50	0.2	0.32	0.05	0.23
f, There are clusters of galaxies as massive as Coma at $z \approx 1$	0.8	0.05	0.14	0.4	0.42	0.95	0.86	0.95	0.86	0.95	0.86
4. The largest structures											
a, The galaxy distribution tends to be sheet-like	1.0	0.6	0.60	0.95	0.95	0.8	0.80	0.95	0.95	0.05	0.05
b, There are coherent structures in the galaxy distribution $> 50 h^{-1}$ Mpc across	1.0	0.1	0.10	0.2	0.20	0.95	0.95	0.05	0.05	0.1	0.10
c, Void diameters can exceed $25 h^{-1}$ Mpc	1.0	0.95	0.95	0.95	0.95	0.05	0.05	0.1	0.10	0.9	0.90
d, Mean separation of rich clusters is $\sim 60 h^{-1}$ Mpc	1.0	0.95	0.95	0.9	0.90	0.5	0.50	0.1	0.10	0.8	0.80
e, Cluster clustering length is $22(\pm 4) h^{-1}$ Mpc	0.7	0.05	0.18	0.8	0.71	0.5	0.50	0.4	0.43	0.9	0.78
f, The richest 3% of Abell clusters have mass $> 10^{15} h^{-1} M_{\odot}$, central $M/L \approx 300 h$	0.9	0.1	0.14	0.95	0.90	0.5	0.50	0.5	0.50	0.9	0.86
g, The r.m.s. peculiar velocity smoothed over $r = 30 h^{-1}$ Mpc is $\geq 500 \text{ km s}^{-1}$	0.7	0.05	0.18	0.7	0.64	0.6	0.57	0.05	0.18	0.95	0.82
5. Small-scale clustering											
a, The r.m.s. relative velocity of galaxies at $300 h^{-1}$ kpc separation is $\sim 300 \text{ km s}^{-1}$	0.9	0.5	0.50	0.2	0.23	0.2	0.23	0.05	0.10	0.9	0.86
b, Galaxy clustering length is $r_0 = 5.5 \pm 1$ Mpc	1.0	0.95	0.95	0.05	0.05	0.95	0.95	0.5	0.50	0.95	0.95
c, At $r \leq r_0$ the galaxy distribution approximates a fractal with $D = 1.23$	1.0	0.95	0.95	0.1	0.10	0.5	0.50	0.8	0.80	0.8	0.80
d, The power law ξ_{gg} breaks at $r \approx 2r_0$	1.0	0.1	0.10	0.5	0.50	0.1	0.10	0.5	0.50	0.05	0.05
e, Early type galaxies prefer dense regions	1.0	0.6	0.60	0.5	0.50	0.5	0.50	0.5	0.50	0.5	0.50
f, Spiral galaxies formed outside protoclusters	1.0	0.95	0.95	0.05	0.05	0.5	0.50	0.3	0.30	0.95	0.95
g, Dwarfs avoid the voids defined by giants	0.8	0.05	0.14	0.9	0.82	0.05	0.14	0.3	0.34	0.95	0.86
6. Evolution of galaxies at $z < 1$											
a, There are young galaxies at low z	0.8	0.95	0.86	0.95	0.86	0.2	0.26	0.1	0.18	0.05	0.14
b, The comoving number density of bright galaxies is unchanged ($\pm 50\%$) since $z \approx 0.7$	0.9	0.05	0.10	0.05	0.10	0.5	0.50	0.5	0.50	0.95	0.90
c, Spiral disks are not accreting massive objects	0.9	0.1	0.14	0.2	0.23	0.5	0.50	0.7	0.68	0.8	0.77
d, The age difference between disk and spheroid populations in a spiral galaxy ≥ 3 Gyr	0.7	0.9	0.78	0.5	0.50	0.8	0.71	0.5	0.50	0.3	0.36
7. Structure of galaxies											
a, The frequency distribution of the circular velocity v_c has a hard upper cutoff	1.0	0.2	0.20	0.2	0.20	0.2	0.20	0.5	0.50	0.5	0.50
b, At $hr \approx 1$ to 30 kpc, v_c is approximately flat	1.0	0.9	0.90	0.1	0.10	0.1	0.10	0.2	0.20	0.7	0.70
c, v_c correlates well with luminosity	1.0	0.9	0.90	0.6	0.60	0.5	0.50	0.6	0.60	0.5	0.50
d, The dark matter in haloes is baryonic	0.5	0.05	0.28	0.05	0.28	0.3	0.40	0.5	0.50	0.95	0.72
e, Disk alignment correlates poorly with surrounding galaxies	1.0	0.5	0.50	0.1	0.10	0.5	0.50	0.1	0.10	0.5	0.50
f, Elliptical alignment correlates with environment	1.0	0.7	0.70	0.9	0.90	0.6	0.60	0.9	0.90	0.3	0.30
g, Ellipticals have radial colour gradients	1.0	0.95	0.95	0.9	0.90	0.5	0.50	0.9	0.90	0.05	0.05
h, Metal-poor stars are rare in disks	0.8	0.5	0.50	0.1	0.18	0.05	0.14	0.2	0.26	0.8	0.74
i, Globular clusters cluster around galaxies	1.0	0.8	0.80	0.5	0.50	0.5	0.50	0.5	0.50	0.05	0.05
$\log_{10} \Pi r_i$ $P_n (< \Pi r_i)$		-14.7 0.74		-21.8 0.03		-15.9 0.56		-17.3 0.37		-14.8 0.72	

formation, and we see evidence of these effects on larger scales, for example in the stripping of gas from galaxies in clusters of galaxies. The XPL theory considers the possibility that non-gravitational processes had a significant role in determining the character of the mass distribution on even larger scales.

Baryonic dark matter (BDM) is necessarily accompanied by primeval isocurvature fluctuations (where the primeval mass distribution is uniform, and the entropy per baryon is a random function of position)⁹. This is the only viable low- Ω universe in which the principal constituents are baryons and electromagnetic radiation, the existence of which is indisputable. In this theory, the bulk of galaxy masses were assembled at large redshifts, $z \geq 10$, in marked contrast to the CDM theory, which predicts that galaxies were assembled fairly recently.

We confine this survey to the above five models, which are the most commonly discussed. There is also a variety of hybrid models in which, perhaps out of a sense of desperation, elements from different models are combined (such as hot and cold dark matter), and there are models that introduce exotic new concepts such as late-time phase transitions which produce domain walls in the Universe at recent epochs, and cosmic textures that untangle late enough to seed large-scale structure formation¹⁰. It is too soon to decide whether any of these models should be included in the cosmic book of odds.

Phenomena

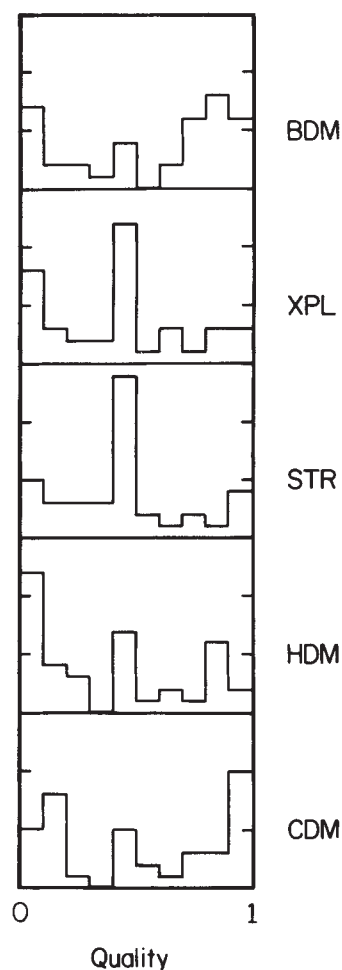
For inclusion in Table 1 we have attempted to select the phenomena that observers consider reasonably well established and which theorists use (or which we think they ought to use) in developing models of how galaxies formed. The first entry in each row indicates the phenomenon we have in mind. The second entry is our estimate of the weight w appropriate for the phenomenon. A weight near unity means the effect is well

established. A weight near zero means either that one can find reasonable interpretations of the observational evidence for and against the reality of the effect, or that the evidence either way does not seem very strong. In some cases, lack of reality of the effect probably implies reality for something else. For example, it is thought that the likely possibilities are that the density parameter Ω is very close to unity or that it is within a factor of two of 0.1. We have entered the latter in line 1a, rather than the theorists' preference, $\Omega = 1$, because $\Omega \approx 0.1$ is the straightforward interpretation of the bulk of the observational evidence. (Dynamical mass estimates generally indicate $\Omega \approx 0.1$ (ref. 2). Also, lowering Ω increases the volume of space, which helps account for the large numbers of objects seen in deep galaxy counts; this explanation is not unique but it does have the virtue of simplicity¹¹.) On the other hand, the theoretical argument for $\Omega = 1$ is of a character that has been successful in other cases despite apparently low odds, so we adopt a fairly modest weight for $\Omega \approx 0.1$.

The first column under each model is our estimate of the probability p that the picture can account for the effect, if real. The second column is a quality factor that convolves the probability p and the weight w for the phenomenon using equation (1) below. Because theorists are ingenious, we have set the minimum value of the probability at $p = 0.05$, no matter how unlikely it seems that the theory can account for the effect. In some cases, our estimate of p would be close to the community consensus. There are other cases in which some would reach a very different resolution of the conflicting lines of argument. Perhaps the situation is best illustrated by some examples of our considerations.

Where we can see no good argument for deciding either way, we assign $p \approx 0.5$. This applies to some obviously relevant and well documented phenomena, such as the tendency for early

FIG. 1 Frequency distributions of quality ratings, equation (1), for the five theories in Table 1.



type elliptical and S0 galaxies to prefer dense neighbourhoods, and spirals the less dense regions (line 5e in the table). The effect is often discussed in connection with theories of galaxy formation, but so far only the CDM model has provided calculations that give objective reasons to claim predictions, rather than those based on the benefit of hindsight. Where the advocates of a theory have concentrated almost exclusively on one possibility, either reality or non-reality of an effect, we assume that that possibility is judged to be the more likely requirement of the theory, and have assigned p accordingly. Therefore, as most discussions of galaxy formation by cosmic strings assume that the Universe is dominated by non-baryonic dark matter, we have assigned a relatively low p for this picture if Ω proves to be small. Of course, the odds would change if it were shown that galaxy formation by strings in a baryon-dominated universe were promising.

In several cases, a theory agrees well with an observation because it was so designed. The large values of p in such cases testify mainly to the ingenuity of the theorists, but it is still relevant to know that the theory can be adjusted to fit the observations. Thus, for example, the successful prediction in CDM theory of the abundance of galaxy concentrations comparable to great clusters⁵ is not entirely unexpected because the theory is normalized to the observed r.m.s. fluctuation in galaxy counts on a scale comparable to the clusters, but the fact that this adjustment is possible is significant (line 4d in Table 1). Nevertheless, it is disconcerting that, with a bias factor of two, the CDM theory seems to underpredict the abundance of the richest known clusters¹² (line 4f). One resolution is to lower the bias, despite the fact that a somewhat higher bias seems to be needed to reconcile the low relative velocities of galaxies with the large mean mass per galaxy implied by the assumption $\Omega = 1$ (line 5a). Lower bias also is the preferred means of reconciling CDM theory with the large-scale bulk flow data¹³ (line 4g). It is, however, hardly admissible to let the degree of biasing be a function of the phenomenon to be explained, despite the long historical tradition in science of 'saving the phenomena'.

Dark matter. If Ω were significantly less than unity, the motivation for postulating exotic matter in the CDM and HDM would be eliminated, so in line 1a they are assigned low p for this eventuality. A critical observation would be the detection of exotic dark matter particles from the massive halo of our Galaxy¹⁴. We assign somewhat greater weight to the possibility that the dark matter in haloes is baryonic rather than exotic (line 7d), because haloes do contain stars; at least one dark very-low-mass 'star' (Jupiter) is known, and such stars in haloes would be good places for the baryons needed for light-element nucleosynthesis³. It is a potential source of concern that, despite intensive searches, we have no firm evidence for a single brown dwarf (<70 Jupiter masses) outside the Solar System; on the other hand, all published unsuccessful searches may have been looking in the wrong places by seeking binary companions to main-sequence stars. The most logical BDM candidates are objects such as brown dwarfs, white dwarfs and neutron stars, all of which are known to exist in considerable numbers, and are likely to have been produced in large numbers (at least in the latter two cases) during the earliest epoch of protogalactic star formation. Perhaps it is time to reassess some of the arguments that purport to eliminate these options¹⁵.

The high degree of the isotropy of the thermal cosmic background radiation (CBR)¹⁶ is a problem only for the HDM model (and for the now generally abandoned model of galaxy formation out of primeval adiabatic density fluctuations in a baryon-dominated universe¹⁷). In the HDM model, the large primeval mass coherence length requires that the amplitude of the primeval mass density fluctuations be relatively large to get nonlinear structures to form by the present epoch. The high amplitude tends to overproduce small-angular-scale anisotropy in the CBR (line 2b), and there is no straightforward recourse

in HDM theory to smoothing such anisotropies by early re-ionization.

Before the launch of the Cosmic Background Explorer (COBE), there were indications of a distortion in the short-wavelength part of the CBR spectrum, roughly similar to what would be produced if a considerable amount of energy were added by Compton-Thomson scattering from a hot plasma. This effect would be expected in an explosion model because it is difficult to see how sufficient energy to push matter the distances needed to produce the observed large-scale structures (lines 4b–d) could be applied without perturbing the spectrum. But the COBE results show that the distortion is not present, and that the CBR spectrum is very close to thermal¹⁸, important evidence against the explosion model.

Quasars are observed at $z \approx 5$ (line 3a), and the absorption lines in the spectra of these distant objects indicate that at this epoch the intergalactic medium was either already highly inhomogeneous, the baryons strongly concentrated into compact objects, or that the intergalactic medium was strongly ionized¹⁹ (line 3d). Moreover, the heavy elements seen in quasar absorption lines having higher column densities require that much of the baryonic matter in the Universe was enriched to a level of the order of one per cent of the solar abundance. This is a problem for the HDM theory because star formation cannot commence until the first 'pancakes' collapse, and we can see no way for that to happen at $z \approx 5$ without greatly overproducing large-scale clustering at the present epoch. In the CDM theory, there are small-scale density fluctuations that can initiate star formation as early as $z \approx 10$, and possibly ionize the intergalactic medium. If a protoquasar is assumed to be associated with a massive protogalaxy, its assembly at $z > 5$ requires an optimistic assessment of the growth of mass concentrations in the CDM theory, and production of such a quasar would require an impressive efficiency in gas transfer from large radii in the protoquasar to the central engine during the short time available²⁰. That is why we assign odds of about one in twenty to quasar formation at $z \approx 5$ in CDM theory (line 3a). Of course, the engine of a quasar could be a condensate of relatively low mass whose collapse preceded formation of the host galaxy²¹, which might have been captured by a merger or formed in a period of intense infall.

Numerical simulations indicate that in the CDM theory most of the objects that one might identify with haloes of galaxies at the present epoch consisted of many well separated fragments at $z \approx 1$ (see Fig. 2 in ref. 4). This does not fit with the evidence (from counts of galaxies as a function of redshift²²) that the abundance of objects with the luminosity of a giant galaxy has changed less than a factor of two since $z \approx 0.5$ (line 6b), or that the disks of spiral galaxies have not been disturbed by the rain of debris to be expected from late galaxy assembly (line 6c).

The masses of galaxies, as measured by the circular velocities v_c of stars and gas at a fixed distance from the centre, are highly variable, but there is a very sharp upper cutoff, at v_c about twice the value in our Galaxy²³. Because one might expect that the merging of halo fragments in the CDM model would lead to a considerable variation in present halo masses, we assign the theory a rather low probability for accounting for this effect (line 7a). We assign a similarly low probability to the STR model for this because we have seen no arguments explaining how this mass cutoff could arise out of the scale-invariant spectrum of fluctuations expected from a primeval string network. The BDM theory meets with a similar problem: the scale of clustering is not naturally associated with v_c . A related observation that has not been adequately explained is the Fisher-Tully relation, the tight correlation between peak rotational velocity and luminosity, especially in the near infrared (line 7c). Perhaps the answer lies in the non-gravitational dissipation involved in galaxy formation. In particular, cooling sets in below a mass scale of $\sim 10^{12} M_\odot$ (depending on the density in the

protocluster), and cooling certainly is a prerequisite for star formation^{24,25}.

Young galaxies. Because the theories offer very different views of what the Universe might have been like at $z \approx 2-3$, the observations of objects at these redshifts are of considerable interest. As the quasar phenomenon plausibly is associated with youth, the peak in the abundance of quasars at $z \approx 2.5$ (line 3c)²⁶ indicates that this epoch coincides with the peak production of massive galaxies. Also, the remarkably high effective cross-section at $z \approx 2$ of damped Lyman- α (Wolfe) clouds (seen in the spectra of quasars as objects covering $\sim 30\%$ of the sky and having the mass per unit area expected for gas-rich young galaxies) could be interpreted to mean that galaxies were being assembled then²⁷. There are, however alternative explanations of the Wolfe clouds: the low star formation rates measured from detection, and lack of detection, of Lyman- α emission, are characteristic of gas-rich dwarfs or of low-surface-brightness giants (such as Malin 1) rather than normal galaxies, and such objects may indeed dominate the effective cross-section of gas disks producing systems with damped Lyman- α ²⁸.

The highly luminous galaxies discovered as radio sources at $z \approx 2-3$ may be galaxies in the process of forming. They have many of the properties that one would expect to find in galaxies of age ~ 0.3 Gyr, significantly less than the age of the Universe at that epoch (line 3e). They tend to be extended, with irregular shapes and high internal velocities. The optical continuum, presumed to be starlight, and the gas that produces strong emission lines tend to line up with the radio image, an effect that would not be expected to persist for more than a few crossing times, considerably less than the age of the Universe²⁹. On the face of it, this is strong evidence for the presence of newly formed galaxies at $z \approx 2-3$. The extended infrared emission, which also tends to be aligned along the radio jet, can be interpreted in terms of a young stellar population²⁰, notwithstanding earlier claims that this favoured a population of age $\sim 1-3$ Gyr. But there is remarkably little scatter in the observed infrared luminosities of these galaxies, and the infrared luminosities at high redshift fit a smooth extrapolation of the luminosities of radio galaxies observed at low redshifts. One might reasonably expect considerable variability in galaxy luminosities if these objects were in the process of formation. The simpler (but not unique) interpretation would be that the stellar populations have reached a mature state, which would require formation of the stars at high redshift³¹. A further complication is that these objects are not necessarily characteristic of normal galaxies: they were discovered because they are radio sources. Indeed, it has been argued that the optical luminosities of these objects were triggered by radio-jet interactions with protogalactic gas, a phenomenon that may or may not be distinct from the way galaxies themselves formed, perhaps by conventional gravitational collapse or some other merger-driven process³²⁻³⁴.

Our impression is that the evidence, although hardly compelling, tends to favour the conclusion that a class of galaxies was forming at $z \approx 2-3$. If so, this is a serious problem for the BDM theory but somewhat less serious for the XPL theory because it is easy to imagine some delay between the explosion and the fragmentation of the resulting ridge. It is not clear whether such early galaxy formation is consistent with the CDM theory.

At low redshifts, dwarfs with high gas masses, low abundances of heavy elements, and containing few old-looking stars could be young galaxies³⁵. This would be very unlikely in the BDM model (line 6a), although there still is ample time for the recycling of gas through stars, leading to bursts of star formation in young-looking galaxies that can have been in existence for a long time.

Sheets and clusters. One of the more dramatic recent events in cosmology was the confirmation that galaxies tend to lie on sheet-like arrangements. The Giovanelli-Haynes Perseus-Pisces supercluster is at least $Hr = 5,000 \text{ km s}^{-1} \text{ long}$ ³⁶, and the 'Great

Wall' seen in the SLICE survey is even wider³⁷. Equally dramatic was the prediction of structures of this type, albeit on smaller scales, in theories such as HDM that involve pancake collapse, and in the ridges naturally produced by explosions. But producing such structures in explosions, while leaving the fragile-looking loosely clustered arrangement of galaxies around the sheets, is a considerable challenge, as we have indicated in line 4b. It is also difficult to see how explosions could produce the broad correlation length of the peculiar velocity field, as indicated in line 4g. We have assigned this latter phenomenon a relatively low weight, however, because of the possibility that it is an artefact, perhaps caused by spatial variations of intrinsic galaxy parameters³⁸.

Can large-scale sheet-like patterns develop in models, such as CDM and BDM, where structure develops in a hierarchical way from small to large scale? The situation is different to that in the HDM model, where the pancake collapse is produced at the mass coherence length. In CDM and BDM theories, there is no mass coherence length. The expected result is that there is no coherence length in the relative velocity field, and so structure develops by hierarchical clustering rather than the pancake process. If large-scale sheets like the Great Wall do form in the CDM or BDM models, as suggested by some numerical experiments³⁹, it will be a remarkable example of pattern development out of noise. Pending further discussion of how this could happen, we assign these theories a low rating on this score.

The largest reliably established structures are of particular interest for the CDM theory, because it predicts that the primeval mass distribution is anticorrelated at separations larger than $Hr \approx 1,500 h^{-1} \text{ km s}^{-1}$, and for explosion theories, because the length scale helps fix the energy requirements. The length r_c at which the rich-cluster correlation function is unity (line 4e) is subject to systematic error because of the way in which the clusters were discovered. We are impressed, however, by the consistency of results from samples and estimators that might be expected to have been affected by systematic errors in rather different ways. This includes the use of angular distributions⁴⁰, redshift samples⁴¹, and an X-ray selected sample⁴². If this value is right, it is uncomfortably close to the anticorrelation length in the CDM theory, at which one would not expect to see a positive correlation of positions of objects.

One notable property of the great clusters of galaxies is mentioned in line 4f. Minimum masses for a few cases, such as the Coma cluster, are reliably known to reach $10^{15} h^{-1} M_\odot$ (ref. 43). In the CDM model with the conventional parameters it is difficult to see how this much mass could be assembled early enough to convert the cluster members to early types¹². A related issue is that the mass-to-light ratio in the central parts of a cluster is $\sim 20\%$ of the global mean for an Einstein-de Sitter model. In the CDM theory, this is reasonably accounted for by biasing⁴. In STR theories, however, clusters gravitationally accrete around string loops or wakes, which would be some generations removed from the progenitors of galaxies. For this model it is difficult to understand how clusters could be biased against the collection of mass over light. We have not assigned a very low probability to the STR model for this effect because of the possibility, not yet adequately explored, that the cosmic-string model could be made consistent with a pure baryon universe.

At separations $r \geq 10 h^{-1} \text{ Mpc}$, the galaxy two-point correlation function $\xi_{gg}(r)$ breaks from the power law observed at smaller r (line 5d)⁴⁴. Numerical simulations of the evolution of the mass distribution indicate that it will not be easy to account for this in the BDM theory⁴⁶ (where the slope of ξ_{gg} tends to be much too steep at $\xi_{gg} \approx 1$).

The power-law form of ξ_{gg} on smaller scales, and the behaviour of the three- and four-point correlation functions, are consistent with the description of the small-scale galaxy distribution as a scale-invariant clustering hierarchy, or fractal, with

dimension $D = 1.23$. The production of this pattern has only been successfully demonstrated in the CDM theory⁴⁷, as indicated in line 5c.

Biasing. In CDM and HDM theories, the low dynamical estimates of the density parameter Ω are reconciled with the Einstein-de Sitter model by the biasing assumption that galaxies form with high efficiency in regions of high primeval density, the bulk of the mass remaining in the low-density interstices where it would not be detected in observations of the dynamics of groups and clusters of galaxies. What became of the potential sites of galaxy formation in the low-density regions? It would be reasonable to suspect that many sites end up as irregular dwarfs or unusual-looking galaxies. The observations relevant to this expectation are still being debated, but there is one straightforward case, the distribution of galaxies in our immediate neighbourhood ($Hr \leq 400 \text{ km s}^{-1}$). About 200 galaxies are known in this region, the large number being due to the fact that galaxies of even very low luminosity are visible this close to us. Most of the nearby galaxies are concentrated in a plane coincident with the plane of the Local Supercluster⁴⁸, but there are a few giant spiral galaxies, such as NGC6946, that are not on the plane. Casual inspection shows no gross peculiarities in NGC6946⁴⁹. In the biasing picture, galaxies such as this one would have to have formed in places where the chances of survival were low, and it would seem surprising that they show no evidence of their traumatic youth. As for the absence of dwarf irregular galaxies in the voids, there is, of course, the possibility that such 'galaxies' formed stars very inefficiently, so that they have low surface brightness and are therefore hard to detect—the 'protogalaxy' of Giovanelli and Haynes⁵⁰ provides an example of how difficult it can be to detect such objects. Nor need they be dwarfs—consider the retarded giant Malin 1 (ref. 51), for example. On the other hand, the 'dark' galaxy DDO154 is about two magnitudes fainter than the Small Magellanic Cloud, yet it has an extended dark halo. Carignan and Freeman⁵² point out that galaxies like this could provide the 'dark mass' needed to account for the dynamics of systems of galaxies. DDO154 is in the concentration of galaxies in the plane of the Local Supercluster, however, not in one of the voids where we want to put most of the mass of an $\Omega = 1$ universe. Because DDO154 was discovered and catalogued over two decades ago, and the surveys reveal few candidates for similar objects outside the local concentration, it is reasonable to infer that there are not many analogues in the local voids. On the other hand, the search for galaxies of extremely low surface brightness, especially dwarfs but not excluding giants, has not been sufficiently systematic (focusing only on a few nearby galaxy clusters) to be statistically meaningful.

Because the situation in our neighbourhood is so contrary to what the biasing picture in CDM theory would suggest, we assign the theory a low probability (line 5g). We assign the STR theory a low probability because it seems an unlikely coincidence that the progenitors of dwarf galaxies would place them close to the seeds of giants. We assign the XPL model a moderately low probability because it is easy to imagine that early generations of explosions would leave remnants in the voids produced by later explosions. The BDM model succeeds, because it postulates a low-density universe: there is no excess mass to hide outside groups and clusters of galaxies.

None of the theories explain galaxy structure very satisfactorily. The disk-halo conspiracy (line 7b) is that the rotation speed $v_c(r)$ (of particles moving in a stable circular orbit at radius r) is close to flat from the inner part of the galaxy, where the mass is dominated by stars, to large radii, where the mass is dominated by dark matter⁵³. This seems rather surprising in non-baryonic dark matter theories, although the CDM theory can explain it by the dissipative collapse of the baryonic component. One has a qualitative understanding of the observed gradients in colours of the spheroidal populations of galaxies, as a result of dissipative contraction⁵⁴. The effect is thought to

be less easy to understand if star formation is not contemporaneous with galaxy formation. In the BDM theory, spheroids would form at high density, leading to the possibility of a rapid collapse that erases the gradient (line 7g). The distribution of globular clusters, notably the correlation of colour and age with position in the galaxy, suggest a dissipative origin requiring a gas-rich protogalaxy when such objects formed, expected in the CDM model but not in the BDM (line 7i).

Quality ratings

To quantify the successes and failures of the theories, we convolve the probability p for the theory with the weight w for the phenomenon by the expression

$$r = \frac{1 + 2wp - w}{2} \quad (1)$$

This quality rating has the desired character of a probability. If the odds that a theory accounts for a given effect are about even, $p = 0.5$, then $r = 0.5$, independent of the weight of the phenomenon, as desired. If the weight is close to zero then again $r \approx 0.5$, independent of what the theory predicts. And finally if the weight is high, $w \approx 1$, then $r \approx p$ as desired.

We combine the r_i for the 38 phenomena in the table by taking their product, $\prod r_i$. The logarithm of this product can be compared to the familiar χ^2 statistic. If the r_i were uniformly distributed between zero and unity and were statistically independent, the probability that the product of the r_i is less than R would be

$$P_n(\prod r_i < R) = R \sum_{m=0}^{n-1} \frac{|\log R|^m}{m!} \quad (2)$$

An observationally improbable theory would tend to have significant numbers of small r_i , which would make P_n well below unity, whereas a theory whose predictions are positively correlated with observations would tend to have r_i close to unity, making P_n close to unity.

The values of P_n are listed in the last row of the table. We obtain a low probability for HDM, whereas the probabilities are near 50% for the others. BDM and CDM do emerge, at first glance, somewhat ahead of the pack, but the difference is barely significant. More insight into the relative merits of the various theories can be gleaned from Fig. 1, which shows the frequency distributions of the quality ratings r_i . A promising theory would show a distinct concentration of values toward the right-hand side in this distribution. The HDM theory shows a concentration toward small r_i , consistent with its general lack of popularity, whereas the other four theories show no very convincing signs of correlation between theory and observation.

The bimodal distribution of quality ratings for the CDM theory, with peaks at high and low r_i , reflects its standing in the community: some stress the successes of this theory, others its failures. The BDM theory has been less thoroughly discussed but has a similar distribution of qualities. The quality ratings for the STR and XPL theories have a strong peak at $r \approx 0.5$. This is the result of recent developments, in the former theory the change from dominance by string loops to long strings, in the latter the lack of detection of the perturbation to the spectrum of the cosmic background radiation. Because of these recent changes, which tend to nullify some of the most quantitative predictions of these models, it has been difficult to get a sense of what these theories might imply.

Conclusions

As with our assignments of the 190 quality ratings in Table 1, some of our inferences are fair measures of the general feeling among cosmologists, whereas others may be somewhat controversial. With that cautionary note, we list our conclusions as follows.

■ It can happen, on rare occasions, that a crucial observation convincingly rules out a theory in cosmology. Such was the fate

of the steady-state theory on the discovery of the cosmic background radiation. In general, however, it is difficult to demonstrate convincingly that a theory is wrong because the observations are usually subject to uncertainties of fact or interpretation that its proponents are entitled to stress. Thus, despite the fact that the CDM theory has quality ratings $r_i < 0.2$ in 13 of the 38 phenomena, there is considerable interest in this theory. Our guess is that the strongest motivation for a general loss of interest would be the proposition of a manifestly less problematic theory, but no current theory fits that description.

■ A dramatic observational success need not be significant evidence of the correctness of a theory: it may be accidental or contrived, as must be the case for most of the successes recorded in the table (because only one of the theories can be right). Although a crucial discovery, such as the detection of exotic dark matter particles, could markedly narrow the range of possibilities, it would not be good strategy to depend on such an occurrence. One should demand that a promising candidate be consistent with the full range of critical phenomena. Because that has not happened for any of the five theories as they are now understood, the observational successes that can be claimed for each do not make a strong case for their correctness. Therefore we would not give very high odds that any of these theories is a useful approximation of how galaxies were actually formed.

■ In contrast to the dismal theoretical situation, we are very impressed with the number of phenomena that can be entered in the table. A survey 10 years ago would have yielded a considerably shorter list, and the constraints would have been much less interesting because implications of different phenomena interact. In the next few years the weights may change and a few entries may be dropped (as happened during the preparation of this review) but that will probably be more than compensated by additions. Some can be anticipated, such as the new measures of velocity fields from large-scale redshift surveys, and surprises are also likely. We also expect that further

exploration of the phenomena on our list will considerably strengthen the observational basis for theories of galaxy formation. Of central importance to cosmology is the mean mass density: if $\Omega = 1$, can we find a rational explanation of why the mass is so well hidden (lines 1a, 1b, 5g)? The character of the large-scale velocity field (line 4g) and of the largest physical structures (line 4b) is a critical test for some theories and of particular interest because, being close to the regime of linear perturbation theory, these phenomena perhaps can be interpreted in terms of primeval conditions in a reasonably convincing way. Observations of objects at high redshifts give us a glimpse of what the universe was like when it was young, and may give us the chance to see galaxies or clusters of galaxies in the process of forming (class 3 in the table). Observations of nearby objects can be difficult to interpret because initial conditions tend to be erased, but that can be compensated by the wealth of observations. We are hoping that class 7 can be considerably enlarged in the next few years.

■ The theorists' analogue of a crucial measurement is a crucial idea that leads us to a new and profitable line of thinking. The great example in cosmology is Einstein's assumption of a homogeneous world model (in the face of all evidence to the contrary). That, with Einstein's theory of general relativity and a very modest input from the observation that galaxy redshifts tend to be positive, led Lemaître to the discovery of the expanding universe. The five theories we have considered attempt a similar thing on a less grand level: each is based on a bold assumption that has only an indirect relation to observation. Another inspiring example, from geology, was the discovery of continental drift and plate tectonics. It required a grand vision, but it also required painstaking deductions from confusing data. Are we ready for that in cosmology? □

P. J. E. Peebles is at the Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544, USA and Joseph Silk is at the Departments of Astronomy and Physics, University of California, Berkeley, California 94720, USA.

1. Peebles, P. J. E. & Silk, J. *Nature* **335**, 601–606 (1988).
2. Peebles, P. J. E. *Nature* **321**, 1–6 (1986).
3. Yang, J., Turner, M. S., Steigman, G., Schramm, D. N. & Olive, K. A. *Astrophys. J.* **281**, 493–511 (1984).
4. Frenk, C. S., White, S. D. M., Davis, M. & Efstathiou, G. *Astrophys. J.* **327**, 507–525 (1988).
5. Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. *Astrophys. J.* **351**, 10–21 (1990).
6. Zel'dovich, Ya. B., Einasto, J. & Shandarin, S. F. *Nature* **300**, 407–423 (1982).
7. Bennett, D. B. & Bouchet, F. R. *Phys. Rev. Lett.* **63**, 2776–2779 (1989).
8. Weinberg, D. H., Ostriker, J. P. & Dekel, A. *Astrophys. J.* **336**, 9–45 (1989).
9. Peebles, P. J. E. *Nature* **327**, 210–211 (1987).
10. Turok, N. *Phys. Rev. Lett.* **63**, 2625–2628 (1989).
11. Koo, D. C. in *The Epoch of Galaxy Formation* (eds Frenk, C. S. et al.) 71–83 (Kluwer, Dordrecht, 1989).
12. Peebles, P. J. E., Daly, R. A. & Juszkiewicz, R. *Astrophys. J.* **347**, 563–576 (1989).
13. Rubin, V. C. & Coyne, G. V. *Large-Scale Motions in the Universe* (Princeton Univ. Press, 1988).
14. Primack, J., Seckel, D., & Sadoulet, B. A. *Rev. Nucl. Part. Phys.* **38**, 751–799 (1988).
15. Tamana, F., Silk, J., Wood, M. A. & Winget, D. E. *Astrophys. J.* (in the press).
16. Wilkinson, D. T. in *Inner Space/Outer Space* (eds Kolb, E. W. et al.) 126–132 (Univ. Chicago Press, 1986).
17. Silk, J. in *Proc. 14th Texas Symp. Relativistic Astrophysics* (N.Y. Acad. Sci., in the press).
18. Mather, J. C. et al. COBE preprint 90–01 (1990).
19. Schneider, D. P., Schmidt, M. & Gunn, J. E. *Astr. J.* **98**, 1507–1522 (1989).
20. Efstathiou, G. & Rees, M. J. *Mon. Not. R. astr. Soc.* **230**, 5P–11P (1988).
21. Silk, J. & Szalay, A. *Astrophys. J.* **323**, L107–L112 (1987).
22. Loh, E. D. *Astrophys. J.* **329**, 24–37 (1988).
23. Giovanelli, R., Haynes, M. P., Rubin, V. C. & Kent, W. F. *Astrophys. J.* **301**, L7–L11 (1986).
24. Silk, J. *Astrophys. J.* **211**, 638–648 (1977).
25. Rees, M. J. & Ostriker, J. P. *Mon. Not. R. astr. Soc.* **179**, 541–559 (1977).
26. Warren, S. J. & Hewett, P. C. *Rep. Prog. Phys.* (in the press).
27. Wolfe, A. M., Turnshek, D. A., Smith, H. E. & Cohen, R. D. *Astrophys. J. Suppl.* **61**, 249–304 (1986).
28. Impey, C. & Bothun, G. *Astrophys. J.* **341**, 89–104 (1989).
29. Spinrad, H. in *The Epoch of Galaxy Formation* (eds Frenk, C. S. et al.) 39–56 (Kluwer, Dordrecht, 1989).
30. Chambers, K. C. & Charlot, S. *Astrophys. J.* **348**, L1–L4 (1990).
31. Lilly, S. J. in *The Epoch of Galaxy Formation* (eds Frenk, C. S. et al.) 63–69 (Kluwer, Dordrecht, 1989).
32. DeYoung, D. *Astrophys. J.* **342**, L59–L62 (1989).
33. Rees, M. J. *Mon. Not. R. astr. Soc.* **239**, 1P–4P (1989).
34. Daly, R. A. *Astrophys. J.* **355**, 416–426 (1990).
35. Kunth, D. & Sargent, W. L. W. *Astrophys. J.* **300**, 496–499 (1986).
36. Haynes, M. P. & Giovanelli, R. *Astrophys. J.* **306**, L55–L59 (1986).
37. Geller, M. J. & Huchra, J. P. *Science* **246**, 897–903 (1989).
38. Silk, J. *Astrophys. J.* **345**, L1–L4 (1989).
39. Park, C. *Mon. Not. R. astr. Soc.* **242**, 59P–61P (1990).
40. Hauser, M. G. & Peebles, P. J. E. *Astrophys. J.* **185**, 757–785 (1973).
41. Bahcall, N. A. *A. Rev. Astr. Astrophys.* **26**, 631–686 (1988).
42. Lahav, O., Edge, A. C., Fabian, A. C. & Putney, A. *Mon. Not. R. astr. Soc.* **238**, 881–885 (1990).
43. Hughes, J. P. *Astrophys. J.* **337**, 21–33 (1989).
44. Groth, E. J. & Peebles, P. J. E. *Astrophys. J.* **310**, 507–517 (1986).
45. Maddox, S. J., Efstathiou, G., Sutherland, W. J. & Loveday, J. *Mon. Not. R. astr. Soc.* **242**, 43P–47P (1990).
46. Peebles, P. J. E. *Astrophys. J.* **297**, 350–360 (1985).
47. White, S. D. M., Frenk, C. S., Davis, M. & Efstathiou, G. *Astrophys. J.* **313**, 505–516 (1987).
48. Tully, R. B. & Fisher, J. R. *Nearby Galaxies Atlas* (Cambridge University Press, 1987).
49. Sandage, A. & Bedke, J. *Atlas of Galaxies*, plate 11 (NASA, Washington, 1988).
50. Giovanelli, R. & Haynes, M. P. *Astrophys. J.* **346**, L5–L7 (1989).
51. Bothun, G., Impey, C. D., Malin, D. F. & Mould, J. R. *Astr. J.* **94**, 23–29 (1987).
52. Carignan, C. & Freeman, K. C. *Astrophys. J.* **332**, L33–L36 (1988).
53. Bahcall, J. N. & Casertano, S. *Astrophys. J.* **293**, L7–L10 (1985).
54. Silk, J. & Norman, C. A. *Astrophys. J.* **247**, 59–76 (1981).

ACKNOWLEDGEMENTS. We thank R. Juszkiewicz and J. Ostriker for discussions. This work was supported in part by the NSF.

A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif

Andrew H. Sinclair, Philippe Berta*, Mark S. Palmer, J. Ross Hawkins, Beatrice L. Griffiths, Matthijs J. Smith, Jamie W. Foster*, Anna-Maria Frischauf, Robin Lovell-Badge† & Peter N. Goodfellow

Human Molecular Genetics Laboratory, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London, WC2A 3PX, UK

† Laboratory of Eukaryotic Molecular Genetics, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

A search of a 35-kilobase region of the human Y chromosome necessary for male sex determination has resulted in the identification of a new gene. This gene is conserved and Y-specific among a wide range of mammals, and encodes a testis-specific transcript. It shares homology with the mating-type protein, *Mc*, from the fission yeast *Schizosaccharomyces pombe* and a conserved DNA-binding motif present in the nuclear high-mobility-group proteins HMG1 and HMG2. This gene has been termed *SRY* (for sex-determining region Y) and proposed to be a candidate for the elusive testis-determining gene, *TDF*.

THE mammalian Y chromosome plays a crucial part in sex determination: an embryo that inherits a Y chromosome develops as a male whereas an embryo lacking a Y chromosome develops as a female¹. The sex determining gene(s) on the Y chromosome induces testicular development, and subsequent male sexual differentiation is a consequence of the hormonal products of the testis². The Y-encoded testis-determining gene has been named *TDF* (testis-determining factor) in humans and *Tdy* (testis-determining Y chromosome) in mouse. Although it is likely that many different genes are required for both male and female sex determination, understanding the mode of action of *TDF* may provide a general model for the genetic control of developmental decisions in mammals.

Attempts to identify and clone *TDF* have exploited detailed maps of the Y chromosome. Three types of map have been constructed: deletion maps from analysis of the genomes of sex-reversed XX males and XY females^{3,4}; a meiotic map of the pseudoautosomal region which is shared by the X and Y chromosomes^{5,6} and a long-range restriction map linking the first two maps⁷. Most XX males have inherited Y-derived sequences, including *TDF*, by terminal exchange between the X and Y chromosomes^{8,9}. A map constructed on the basis of the Y fragments present in different XX males placed *TDF* in the distal part of the Y chromosome adjacent to the pseudoautosomal region³. The meiotic map of the pseudoautosomal region indicated that *MIC2* was the closest known pseudoautosomal locus to the sex-specific part of the Y chromosome⁵. A long-range restriction map beginning at *MIC2* and extending into the sex-specific region of the X and Y chromosomes identified a CpG-rich island on the Y chromosome which represented a possible candidate gene for *TDF*⁷. This gene was cloned after

a chromosome walk initiated 130-kilobase (kb) proximal to the CpG-rich region, and subsequently named *ZFY* (ref. 10). The sequences present in a particular XX male (LGL203) and absent in an XY female (WHT1013) were used to define the position of *TDF* to within an interval of 140 kb and *ZFY* was isolated from this region. Other evidence consistent with identity between *ZFY* and *TDF* included the finding of *ZFY*-related sequences on the Y chromosome of all eutherian mammals tested; the presence of a *ZFY*-related gene, *Zfy-1*, in *Sxr'* (the smallest part of the mouse Y chromosome known to be sex-determining) and the structure of the *ZFY*-encoded protein, which has many features in common with transcription factors^{10,11}. But there were several unexpected findings: first, *ZFX*—a homologue of *ZFY*—was found on the eutherian X chromosome and shown in humans to escape inactivation^{11,12}, and second, in metatherian mammals (marsupials), *ZFY*-related sequences were found not on the Y or X chromosome, but on the autosomes¹³.

Two recent reports have further questioned the role of *ZFY* in male sex-determination. Koopman *et al.* studied the expression of *Zfy-1* and *Zfy-2*, the mouse homologues of *ZFY*, and found that their expression was linked to germ cells—a cell-type not required for normal male development. Furthermore, in *W^c/W^c* mutant mice, testicular development occurred in the absence of detectable *Zfy-1* and *Zfy-2* expression¹⁴. Palmer *et al.* described four sex-reversed XX individuals that had inherited Y-derived sequences not including *ZFY*. Assuming that these individuals are male because of their Y-derived sequences, this mapped *TDF* to the 60 kb proximal to the pseudoautosomal boundary. This result is inconsistent with the published breakpoint of the XY female (WHT1013) (ref. 10) and formally excludes *ZFY* as a candidate for *TDF* (ref. 15).

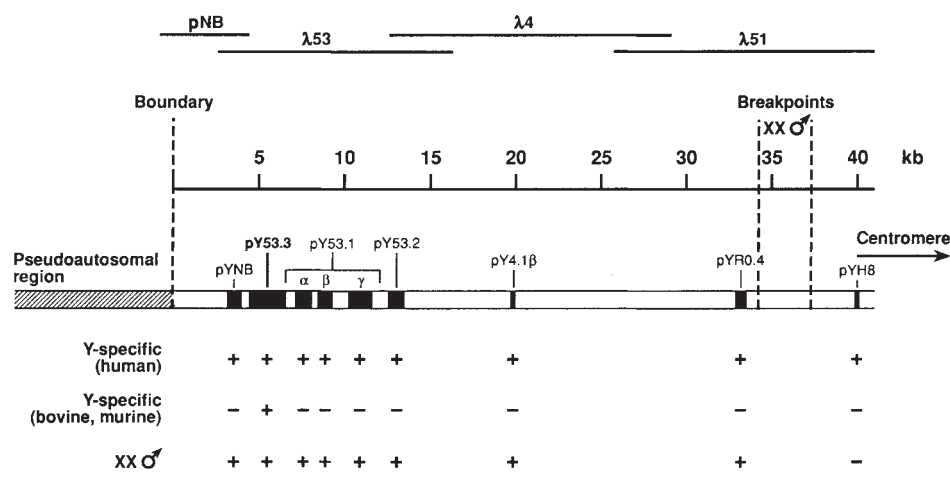
We have tested the hypothesis that the region immediately adjacent to the boundary contains *TDF* by searching for a gene in this region. We have found an open reading frame that is part of a conserved, Y-specific gene. This gene also shares homology with the *Mc* gene of *S. pombe* and with a conserved DNA-binding motif present in non-histone proteins related to HMG1 and HMG2.

Analysis of 35 kb of the Y chromosome

We have previously described a chromosome walk, comprising a series of overlapping λ and cosmid clones, from the pseudoautosomal region, across the pseudoautosomal boundary, to the sex-specific region of the Y-chromosome¹⁶. Probes from this walk were used by Palmer *et al.* on the genomes of XX males to define the region in which *TDF* must lie¹⁵. Since then, three additional probes have been used on the genomes of the same *ZFY*-negative XX males: pY4.1 β , which was positive with the XX males, and pYH8, which was negative (Fig. 1). The third Y-specific probe, pYR0.4, is derived from sequences lying between pY4.1 β and pYH8, and seems to define the break points in the XX males. The probe pYR0.4 detects an 8.5-kb *HindIII* fragment in normal males but only a 4-kb fragment in

* Permanent addresses: Centre de Recherches de Biochimie Macromoléculaire, CNRS LP8402/INSERM U.249, Route de Mende, 34033 Montpellier Cedex, France (P.B.) and Department of Genetics and Human Variation, La Trobe University, Bundoora, Victoria 3083, Australia (J.W.F.).

FIG. 1 Map of distal short arm of human Y chromosome: shaded region (left), pseudoautosomal region; broken line, boundary between pseudoautosomal and Y-specific regions. Top, the three overlapping lambda clones $\lambda 51$, $\lambda 4$, $\lambda 53$ and the plasmid pNB obtained from walking along the Y chromosome. Breakpoints of the XX males, broken lines at 35 kb (see Fig. 2). Black boxes, probes detecting single-copy Y-specific human DNA fragments (+). When these probes were hybridized to male and female DNA from bovine and murine genomes, only pY53.3 detected Y-specific fragments (+). All of the probes except pYH8 hybridized to sequences in the XX males. pY53.3 was known as H2.1 and pY53.2 was called P0.9 (ref. 16).



two related individuals, an XX male and an hermaphrodite; a 6-kb fragment was detected in a third XX male (Fig. 2). The breakpoints in the XX males are clustered around a region that is roughly 35 kb proximal to the boundary (Fig. 1). This result implies that *TDF* is located in sex-specific sequences within 35 kb of the pseudoautosomal boundary. Further refinement of the positions of the breakpoints was not possible because of the highly repetitive nature of the sequences between pY4.1β, pYR0.4 and pYH8.

The following strategy was adopted to locate *TDF* within this 35-kb region of the Y chromosome. DNA from this region was subcloned into fragments of about 4 kb in size. Subsequently, each 4-kb fragment was cleaved with frequently cutting restriction enzymes such as *RsaI*, to produce smaller fragments in the size range 500 base pairs (bp) to 1 kb. A total of 50 probes generated in this manner were examined. Each small fragment was radioactively labelled and used to probe Southern blots of DNA from human males and females, murine males and females, bovine males and females and human-hamster somatic cell hybrids containing the human X or the human Y chromosome. All probes were tested with and without prehybridization to total human DNA to suppress the contribution of repetitive sequences¹⁷. Despite this latter precaution, most probes tested failed to detect unique sequences in the genome of humans but reacted with repetitive elements distributed throughout the genome. These repetitive probes frequently also detected repeat sequences in the bovine genome.

Seven probes were found that detect single-copy Y-specific bands in *EcoRI*-digested human DNA (Fig. 1); pYNB (0.7 kb), pY53.3 (2.1 kb), pY53.1α (0.8 kb), pY53.1β (0.8 kb), pY53.1γ

(1.3 kb), pY53.2 (0.9 kb) and pY4.1β (0.2 kb). But of the seven probes, only pY53.3 also reacted with Y-specific bands in the murine and bovine genomes. A 0.9-kb *HincII* subclone of pY53.3 hybridized most strongly to Y-specific fragments in human, murine and bovine genomic DNA. This subclone was used as a probe in subsequent experiments.

Conservation of pY53.3

A 'Noah's Ark' blot containing DNA from male and female pairs of eutherian mammals was hybridized with the 0.9-kb *HincII* subclone of pY53.3 (Fig. 3a).

The sequences detected by this probe are conserved and male-specific in a wide spectrum of eutherian mammals. At low stringency, additional fragments were found that are shared between males and females, suggesting the existence of pY53.3-related X-linked or autosomal sequences. But these fragments were not detected in a human-rodent somatic cell hybrid that retained the human X chromosome as the only human contribution (Fig. 3b).

Unique sequences that are conserved between the Y chromosomes of different mammalian species are very rare, the only other known example being *ZFY*¹⁰. This result is consistent with pY53.3 reacting with functional sequences on the eutherian Y chromosome.

Sequence analysis of Y-unique sequences

The sequence of pY53.3 was determined by primer-walking. A search of the EMBL DNA sequence database failed to find any sequence related to pY53.3. But inspection of the pY53.3 (2.1 kb) sequence revealed two long open reading frames corresponding

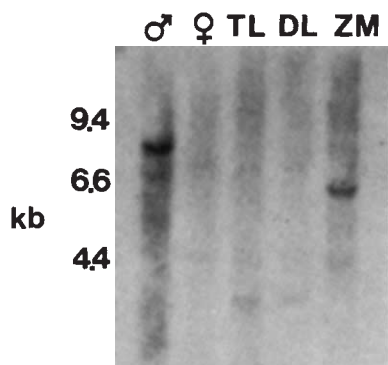


FIG. 2 Southern blot analysis of *HindIII*-digested genomic DNA from: ♂, 46XY cell line, PGF (ref. 28); ♀, 46XX cell line, WT49 (ref. 29); TL, familial cases (XX male); DL (hermaphrodite) and ZM (unrelated XX male) hybridized with pYR0.4. The probe detects an 8.5-kb *HindIII* fragment in normal males but only a 4-kb fragment in the two related individuals TL and DL, and a 6-kb fragment in ZM. This implies that breakpoints in these XX males and hermaphrodite are clustered around a region 35 kb from the boundary (see Fig. 1). DNA size markers (kb) to the left.

METHODS. Genomic DNA (10 µg) was digested with *HindIII*, separated on a 0.8% agarose gel, transferred to Hybond N⁺ (Amersham) and fixed in 0.4 M NaOH (20 min). To suppress repeat sequences present in the probe pYR0.4 it was labelled with ³²P, denatured together with 500 µl (5 mg ml⁻¹) sonicated human placental DNA and prehybridized in 2 ml hybridization buffer at 65 °C for 2 h¹⁷. This probe mixture (2 ml) was added to the filter in a 5 × SSPE buffer, 5 × Denhardt's solution, 0.5% sodium dodecylsulphate, 200 µg ml⁻¹ denatured salmon-sperm DNA and 10% sonicated denatured human placental DNA, and hybridized for 16 h at 65 °C. The filter was extensively washed with 0.2 × SSC buffer, 0.2% SDS at 65 °C and autoradiographed for 6 days.

to 99 and 223 amino acids that overlap in different frames, reading 5'→3' from the centromere towards the pseudoautosomal boundary (Fig. 4a). We screened the PIR protein database using a similarity search algorithm for sequences related to the proteins predicted to be encoded by these open reading frames^{18,19}. The protein putatively encoded by the shorter open reading frame was unrelated to any sequence in the PIR protein database. But the protein putatively encoded by the longer open reading frame was found to have striking similarity both to a portion of the Mc protein encoded by the *mat3-M* of the fission yeast *S. pombe*²⁰ and to a conserved motif in several non-histone proteins related to HMG1 and HMG2^{21,22}. The conserved motif

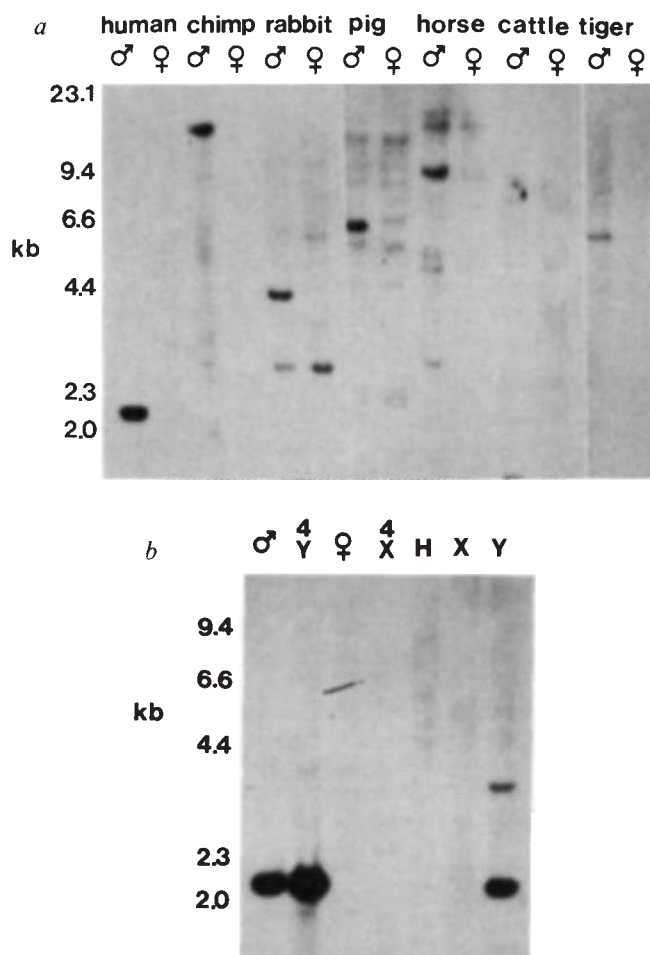


FIG. 3 a, 'Noah's Ark' blot of *HindIII*-digested genomic DNA from male and female pairs of eutherian mammals, hybridized with the 0.9-kb *HincII* fragment of pY53.3. The probe detects male-specific fragments in human (2.1 kb), chimpanzee (~1.8 kb), rabbit (4.2 kb), pig (6.6 kb), horse (~1.0 kb), cattle (1.6 kb) and tiger (6.3 kb). Rabbit ♀ lane is slightly overloaded. b, Southern blot of: ♂, 46XY cell line, PGF (ref. 28); 4Y, 49XYYYY cell line, Oxen (ref. 30); ♀, 46XX cell line, WT49 (ref. 29); 4X, 48XXXX cell line, GM1416B (Coriell Institute for Medical Research, Camden, New Jersey, USA); H, hamster parent cell line, W3GH (ref. 31); X, hamster-human hybrid cell line containing the human X chromosome, CL2D (ref. 32); Y, hamster-human hybrid cell line containing the human Y chromosome, 853 (ref. 33). The filter was hybridized with the (0.9 kb) *HincII* fragment of pY53.3, which detects a 2.1-kb *HindIII* fragment in ♂, 4Y and Y, the intensity of the bands reflects the number of copies of the Y chromosome present. DNA size markers (kb), to the left.

METHODS. Genomic DNA (10 µg) was digested with *HindIII*, separated on a 0.8% agarose gel, transferred to Hybond N⁺ (Amersham) and fixed in 0.4 M NaOH. Probe pY53.3 was labelled with ³²P and added to the filter in 5×SSPE buffer, 5×Denhardt's solution, 0.5% SDS, 10% dextran sulphate, 200 µg ml⁻¹ denatured salmon-sperm DNA, and hybridized for 16 h at 65 °C. The filter was washed in 1×SSC buffer, 0.2% SDS at 65 °C and autoradiographed for 2 days.

covers a stretch of 80 amino acids and suggests a common structural role in these proteins, which could link them functionally (Fig. 4b).

At the 3' end of the conserved motif in pY53.3, the sequence continues for 68 amino acids before reaching a stop codon. No potential splice donor site was found in the DNA sequence of this region but there is a putative polyadenylation signal 133 bp downstream of the stop codon. In the 5' direction there is a potential splice acceptor signal in the pY53.3 DNA sequence at the point where homology between the conserved motif and pY53.3 ends. The open reading frame continues 5' for a distance corresponding to another 75 amino acids and within this region two potential start codons are found in pY53.3.

To test the conservation and hence functional importance of the sequence motifs in pY53.3, the homologous Y-specific rabbit sequence was cloned and sequenced. Within the conserved motif the human and rabbit Y-specific sequences share 64 out of 80 amino acids (80%) with a further eight amino acids showing conservative changes (90% similarity overall) (Fig. 4c). Outside the motif the human and rabbit show only 54% identity. The high degree of homology within the motif strongly suggests pY53.3 contains the coding sequence of a gene.

The other Y-unique probes found in the original search were also sequenced. The pNB and pY53.2 probes did not reveal any open reading frames predicted to encode proteins related to sequences in the EMBL database or the PIR protein database. The probe pY4.1β is part of a larger 1.2-kb *RsaI* fragment that contains an open reading frame predicted to encode a protein related to retroviral reverse transcriptase, commonly found in repetitive sequences. The probe pY53.1 encompasses a 5.6-kb region containing several open reading frames, but none of these was predicted to encode a protein related to sequences in the EMBL database or the PIR protein database. In total, 10.5 kb of the Y chromosome was sequenced in the search for potential coding sequence.

Tissue distribution and expression

A northern blot prepared with poly(A)⁺ RNA from human tissues was hybridized with the 0.9-kb *HincII* fragment of probe pY53.3 (Fig. 5). The probe detects a fragment of about 1.1 kb in adult testis. Bands were not detected in ovary, lung (male) or kidney (male) cell lines, or in male and female lymphoblastoid cell lines. The filter was stripped and re-probed with β-actin to confirm the presence of poly(A)⁺ RNA in the samples (Fig. 5). This result is consistent with a testis-specific transcript being encoded by the pY53.3 Y-specific sequence. This was confirmed using 3' RACE (rapid amplification of complementary DNA ends) polymerase chain reaction (PCR)²³ from adult testis poly(A)⁺ RNA, which showed a poly(A) tract 15 bases downstream from the potential polyadenylation signal, further indicating that this is the 3' end of a Y-specific transcript (data not shown).

Discussion

We have extended the findings of Palmer *et al.*¹⁵ and defined a 35-kb region of Y-specific sequence immediately adjacent to the pseudoautosomal boundary in which *TDF* must reside. An extensive Southern blot analysis of this 35-kb region of the Y chromosome revealed a Y-unique probe, pY53.3, which detects conserved Y-specific sequences in a wide range of eutherian mammals. In related studies, Gubbay *et al.*²⁴ have demonstrated that the equivalent murine sequence is present in *Sxr'* (ref. 25)—the smallest part of the mouse Y chromosome known to be male sex-determining—and is deleted from a mutant Y chromosome that has lost male sex-determining function²⁶. The conservation of pY53.3-related sequences between the Y chromosomes of eutherian mammals suggests that these sequences have a functional role. The location of the pY53.3-related sequences on the Y chromosomes of man and mouse is consistent with a role in male sex determination.

The translated nucleotide sequence of pY53.3 (2.1 kb) contains two open reading frames. The longer open reading frame encodes a region of 120 amino acids shared with the homologous rabbit Y-specific sequence. Within this region there is a conserved motif of 80 amino acids which shows 80% identity, rising to 90% with conservative substitutions. Outside the conserved region the similarity between the residues drops to 54%.

The amino-acid sequence encoded by the longer open reading frame from pY53.3 showed striking similarity to 80 amino acids at the carboxy-terminal of the Mc protein encoded by the *mat3-M* locus of the fission yeast *S. pombe*²⁰. This is the same stretch of 80 amino acids in pY53.3 that is conserved in the rabbit sequence. The mating-type locus, *mat-1*, in the fission yeast has two alternative alleles, *M* and *P*. These alleles are transposed during switch of mating type from either donor loci

mat3-M or *mat2-P* to the *mat-1* locus. Both loci contain two transposable genes (*Mc* and *Mi*, and *Pc* and *Pi*); none of the four genes are related to each other in sequence. The precise function of the four genes is not known; *Mc* and *Pc* are, however, required for mating and all four genes are needed for meiotic competence²⁰. By analogy to the budding yeast it has been suggested that genes of the *mat-1* locus function as transcription factors. This suggestion has been supported by the finding of a diverged homeobox domain in *Pi*²⁰.

The 80-residue conserved motif in pY53.3 also showed homology with a domain found in the nuclear non-histone proteins HMG1 and HMG2. High mobility group (HMG) proteins 1 and 2 are thought to play a part in chromosomal structure and gene activity, and some display enhanced DNA-binding to A + T-rich single-stranded sequences²⁷. HMG1 and HMG2 are

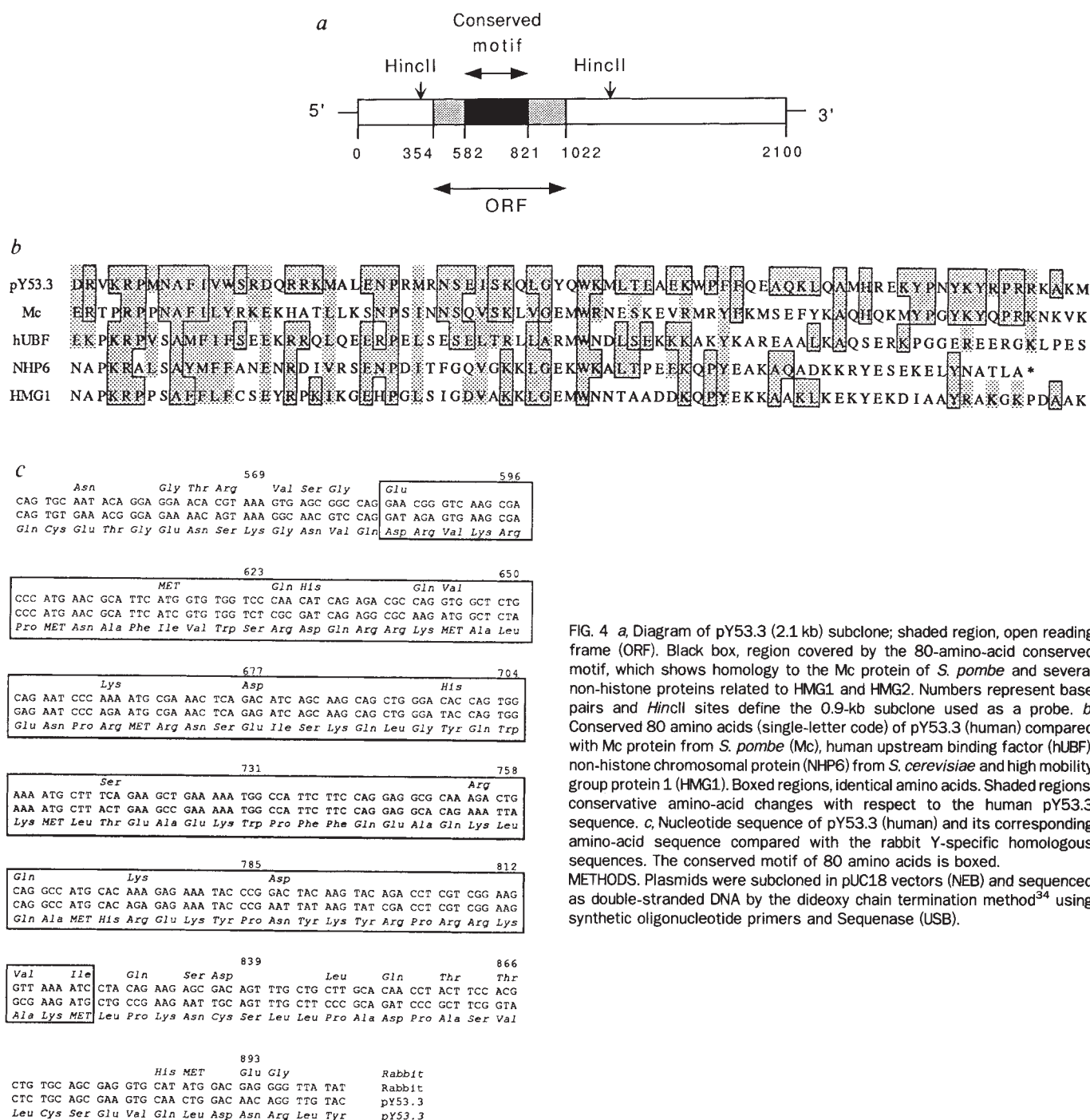


FIG. 4 *a*, Diagram of pY53.3 (2.1 kb) subclone; shaded region, open reading frame (ORF). Black box, region covered by the 80-amino-acid conserved motif, which shows homology to the Mc protein of *S. pombe* and several non-histone proteins related to HMG1 and HMG2. Numbers represent base pairs and *HincII* sites define the 0.9-kb subclone used as a probe. *b*, Conserved 80 amino acids (single-letter code) of pY53.3 (human) compared with Mc protein from *S. pombe* (Mc), human upstream binding factor (hUBF), non-histone chromosomal protein (NHP6) from *S. cerevisiae* and high mobility group protein 1 (HMG1). Boxed regions, identical amino acids. Shaded regions, conservative amino-acid changes with respect to the human pY53.3 sequence. *c*, Nucleotide sequence of pY53.3 (human) and its corresponding amino-acid sequence compared with the rabbit Y-specific homologous sequences. The conserved motif of 80 amino acids is boxed.

METHODS. Plasmids were subcloned in pUC18 vectors (NEB) and sequenced as double-stranded DNA by the dideoxy chain termination method³⁴ using synthetic oligonucleotide primers and Sequenase (USB).

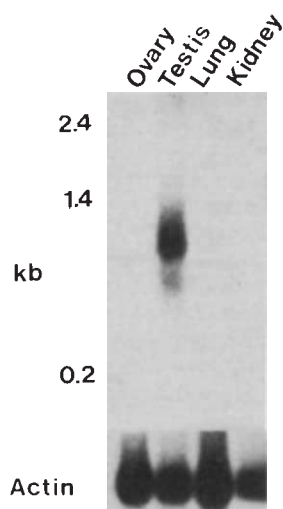


FIG. 5 Northern blot analysis of poly(A)⁺ RNA from ovary, adult testis, lung (male) and kidney (male), hybridized with the 0.9-kb *HincII* fragment of pY53.3. The probe detects a transcript of ~1.1 kb in adult testis and in no other tissue tested. Below is the same filter stripped and re-probed with β -actin, confirming the presence of poly(A)⁺ RNA in all the tissue samples. Poly(A)⁺ RNA was also prepared from three lymphoblastoid cell lines 49XYYY cell line (Oxen)³⁰, 46XY cell line (PGF)²⁸, 46XX cell line (WT49)²⁹, and probed as above, but no transcripts were detected (data not shown). METHODS. RNA was prepared from each tissue as described previously³⁵ and poly(A)⁺ messenger RNA selected by poly(A) tract isolation system (Promega). Poly(A)⁺ RNA (8 μ g) was separated on a 1% agarose gel containing 2.2 M formaldehyde, transferred to Hybond N (Amersham), ultraviolet-irradiated, and hybridized with ³²P-labelled 0.9-kb *HincII* fragment of pY53.3. Hybridization was at 65 °C in 3 $\times$ SSC buffer, 5 $\times$ Denhardt's solution, 200 μ g ml⁻¹ denatured salmon-sperm DNA, 6% polyethylene glycol and 0.1% SDS. The filter was washed in 1 $\times$ SSC buffer, 0.1% SDS at 65 °C and autoradiography was for 6 days at -70 °C. Conditions for re-probing the filter with β -actin were as above but the filter was washed in 0.1 $\times$ SSC buffer, 0.1% SDS at 65 °C and autoradiography was for 8 h.

not known to regulate specific gene sequences but are associated with regions of transcriptionally active chromatin. Within HMG1 and HMG2 there is a motif, the HMG box, which has been found in the non-histone chromosomal protein NHP6 of *Saccharomyces cerevisiae*²², the yeast ARS-binding protein, ABF2, and the human nucleolar transcription factor hUBF (human upstream binding factor)²¹. The hUBF product is an RNA polymerase I transcription factor that interacts with sequence-specific DNA regions. This motif might represent a novel class of DNA-binding domains²¹. The conserved binding motif seems to be present in a large family of related sequences perhaps originating from an early HMG-like nonspecific DNA-binding structure.

It is tempting to speculate that Mc is also a transcription factor and that the conserved motif shared with the human pY53.3 sequence is a DNA-binding domain. The only structural evidence to support this conjecture is the high Arg-Lys content of both Mc and the conserved motif of pY53.3, which is 25% Arg-Lys.

We have termed the human Y-located gene defined by pY53.3 *SRY* (sex-determining region Y). The gene has been defined with respect to its location because only the homology to other genes suggests a DNA-binding function. But the putative nucleic-acid-binding motif within *SRY* and its testis-specific expression are consistent with *SRY* having a role in the developmental regulation of the testis.

The presence of a 3' stop codon and a poly(A) tract in cDNA

from 3' RACE PCR implies that the open reading frame in pY53.3 corresponds to the last exon of *SRY*. At the 5' end of the conserved motif sequence in *SRY* there is a potential splice acceptor site; but as homology between the rabbit and human genomic sequences continues past this site it may not represent an intron-exon boundary. This question will be resolved by the isolation and sequencing of transcripts from the *SRY* gene.

The 35 kb of Y-specific sequences immediately adjacent to the pseudoautosomal boundary are rich in repetitive sequences and this has hampered analysis. The open reading frame in pY53.3 is the only well-conserved sequence detected; but this approach could have failed to detect small exons, especially if they are close to repetitive elements or are only weakly conserved between species. Therefore, we cannot exclude the existence of genes other than *SRY* in the human sex-determining region.

We have described a novel, transcribed gene, *SRY*, present in the sex-determining region of man and mouse. Sequences homologous to *SRY* are located on the Y chromosome in all eutherian mammals tested. *SRY* encodes a protein with a potential DNA-binding domain, which is shared with the Mc protein of the mating-type locus of *S. pombe* and several non-histone proteins related to HMG1 and HMG2. *SRY* is currently the best candidate for *TDF*, the male sex-determining gene in humans. Proof of identity between *SRY* and *TDF* will require mutational analysis of XY females or the production of sex-reversed transgenic mice. □

Received 11 June; accepted 22 June 1990

- Goodfellow, P. N. & Darling, S. M. *Development* **102**, 251-258 (1988).
- Jost, A., Vigier, B., Prepin, J. & Perchellet, J. P. *Recent Prog. Horm. Res.* **29**, 1-41 (1973).
- Vergnaud, G. et al. *Am. J. hum. Genet.* **38**, 109-124 (1986).
- Guellaen, G. et al. *Nature* **307**, 172-173 (1984).
- Goodfellow, P. J., Darling, S. M., Thomas, N. S. & Goodfellow, P. N. *Science* **234**, 740-743 (1986).
- Weissenbach, J., Leveilliers, J., Petit, C., Rouyer, F. & Simmler, M.-C. *Development* **101S**, 67-74 (1987).
- Pritchard, C. A., Goodfellow, P. J. & Goodfellow, P. N. *Nature* **328**, 273-275 (1987).
- Ferguson-Smith, M. A. *Lancet* **ii**, 475-476 (1966).
- Petit, C. et al. *Cell* **49**, 595-602 (1987).
- Page, D. C. et al. *Cell* **51**, 1091-1104 (1987).
- Palmer, M. S., Berta, P., Sinclair, A. H., Pym, B. & Goodfellow, P. N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1681-1685 (1990).
- Schneider-Gädick, A., Beer-Romero, P., Brown, L.G., Nussbaum, R. & Page, D. C. *Cell* **57**, 1247-1258 (1989).
- Sinclair, A. H. et al. *Nature* **336**, 780-782 (1988).
- Koopman, P., Gubbay, J., Collignon, J. & Lovell-Badge, R. *Nature* **342**, 940-942 (1989).
- Palmer, M. S. et al. *Nature* **342**, 937-939 (1989).
- Ellis, N. A. et al. *Nature* **337**, 81-84 (1989).
- Sealey, P. G., Whittaker, P. A. & Southern, E. M. *Nucleic Acids Res.* **13**, 1905-1922 (1985).

- Smith, T. F. & Waterman, M. S. *J. molec. Biol.* **147**, 195-197 (1981).
- Collins, J. F., Coulson, A. F. W. & Lyall, A. *CABIOS* **4**, 67-71 (1988).
- Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. *EMBO J.* **7**, 1537-1547 (1988).
- Jantzen, H.-M., Adman, A., Bell, S. P. & Tjian, R. *Nature* **344**, 830-836 (1990).
- Kolodrubetz, D. & Burgum, A. *J. biol. Chem.* **265**, 3234-3239 (1990).
- Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8998-9002 (1988).
- Gubbay, J. et al. *Nature* **346**, 245-250 (1990).
- McLaren, A., Simpson, E., Tomonari, K., Chandler, P. & Hogg, H. *Nature* **312**, 552-555 (1984).
- Lovell-Badge, R. & Robertson, E. *Development* **109**, 635-646 (1990).
- Wright, J. M. & Dixon, G. H. *Biochemistry* **27**, 576-581 (1988).
- Goodfellow, P. J. et al. *Ann. hum. Genet.* **53**, 15-22 (1989).
- DeKretser, T. A., Crumpton, M. J., Bodmer, J. G. & Bodmer, W. F. *Eur. J. Immun.* **12**, 600-606 (1982).
- Bishop, C. E. et al. *Nature* **303**, 831-832 (1983).
- Westerveld, A., Visser, R. P. L. S., Meera Khan, P. & Bootsma, D. *Nature new Biol.* **234**, 20-24 (1971).
- Goss, S. J. & Harris, H. *Nature* **255**, 680-683 (1975).
- Burk, R. D., Ma, P. & Smith, K. D. *Molec. cell. Biol.* **5**, 576-581 (1985).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156-159 (1987).

ACKNOWLEDGEMENTS. We thank M. Fellous for XX male DNA, I. Pickford for providing tissue samples, I. Goldsmith for preparing oligonucleotides and C. Middlemiss for typing this manuscript. A UK patent is pending on the *SRY* sequence.

A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes

John Gubbay, Jérôme Collignon, Peter Koopman, Blanche Capel, Androulla Economou, Andrea Münsterberg, Nigel Vivian, Peter Goodfellow* & Robin Lovell-Badge

Laboratory of Eukaryotic Molecular Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

* Human Molecular Genetics Laboratory, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

A gene mapping to the sex-determining region of the mouse Y chromosome is deleted in a line of XY female mice mutant for *Tdy*, and is expressed at a stage during male gonadal development consistent with its having a role in testis determination. This gene is a member of a new family of at least five mouse genes, related by an amino-acid motif showing homology to other known or putative DNA-binding domains.

WHETHER an animal develops into a male or a female depends on a switch mechanism¹. This may be environmental, as in alligators, where the temperature at which the embryo is incubated is important²; or chromosomal, as in *Drosophila* or *Caenorhabditis elegans*³, where the activity of genes involved in sexual dimorphism is dependent on the number of X chromosomes relative to sets of autosomes in a cell (the X:autosome ratio). Alternatively, the switch can be genetically controlled, as it is in mammals where a gene on the Y chromosome is responsible for male development⁴⁻⁶. This gene is termed *TDF* in humans and *Tdy* in mice.

In eutherian mammals evidence suggests that sex determination is equivalent to testis determination¹. The gonad is composed of cells derived from four lineages: the supporting cells, steroid-producing cells, connective tissue cells and germ cells. In the developing testis, the differentiation of Sertoli cells from the supporting cell lineage is thought to result in cells of the other lineages following the male pathway⁷. The nature of this influence is not known, however it results in the differentiation of Leydig cells from the steroid-producing cell lineage, the induction of mitotic arrest in the germ cells and the proliferation and organization of connective tissue and blood vessels into the testicular pattern. The Sertoli cells themselves proliferate rapidly and organize into the 'testis cords', apparent in the mouse at about 12.5 days post coitum (d.p.c.), and characteristic of the first signs of male development. In the female pathway, the first noticeable event is the entry of germ cells into meiosis^{8,9}, followed by the differentiation of follicle cells and their aggregation around the oocytes. The cells of the steroid-producing cell lineage give rise to theca cells, and there is little proliferation or organization of connective tissue cells. In the absence of germ cells the ovary fails to develop, which suggests that ovarian determination is dependent on gene action within the germ cells¹⁰. Germ cells are not required for testis development^{11,12}.

The ovarian pathway can be thought of as the normal or 'default' pathway, and the consequence of expression of *Tdy* in the male is to switch the fate of the supporting cell precursors

in the indifferent gonad (genital ridge) from that of follicle cells to that of Sertoli cells. It is assumed that *Tdy* must act within the context of other regulatory molecules that are initially responsible for the formation of the urogenital ridge and then for the development of the gonad itself. Elucidating the involvement of *Tdy* in this process, however, might provide a paradigm of how genes influence developmental fate.

The study of cases of human sex reversal arising from abnormal X-Y interchange at meiosis has led to the precise mapping of *TDF* to a short segment of the Y chromosome adjacent to the pseudoautosomal boundary¹³⁻¹⁶. The analysis of four such XX males¹⁷ placed the gene to within a maximum of 60 kilobases (kb) from the boundary. The *ZFY* gene, which had previously been considered a good candidate for *TDF*¹⁸, lies outside this chromosomal region. A more detailed search for the breakpoints in these XX males shows that the minimal portion of the Y chromosome able to confer maleness extends just 35 kb distal to the boundary¹⁹. It is therefore assumed that *TDF* must lie somewhere within this small region. In the accompanying paper by Sinclair *et al.*¹⁹, genomic clones covering 35 kb of the region adjacent to the pseudoautosomal boundary on the human Y have been used to look for the presence of evolutionarily conserved sequences, with the hope of identifying new candidates for *TDF*. One of these clones detected conserved sequences on the Y chromosome of a wide range of eutherian mammals.

We describe here the cloning of the corresponding sequence from the mouse Y chromosome. The mouse gene contains an open reading frame homologous to that found in the human gene. The predicted protein product of these genes includes a domain characterized by a region of homology with several known or putative DNA-binding proteins, including human upstream binding factor (hUBF; ref. 20) and the Mc protein²¹, the product of one of the mating-type genes of the fission yeast *Schizosaccharomyces pombe*. Four additional genes not linked to the Y chromosome, expressed at 8.5 d.p.c. in mouse, were found by their close homology within this protein domain. Genetic mapping places the Y-specific sequence within the sex-determining region of the mouse Y chromosome. Further, the developmental time and tissue of expression of this gene is consistent with that expected for *Tdy*.

Southern analysis

The sex-reversed mutation *Sxr*, has helped to define the position of *Tdy* in the mouse. *Sxr* probably arose by translocation of the small short arm of the Y chromosome onto the pseudoautosomal region, located at the distal end of the long arm²²⁻²⁵. Through obligatory crossover in the pseudoautosomal region during male meiosis the *Sxr* fragment may be transferred to the X chromosome, and resulting XX *Sxr* offspring are male²³. Apart from *Tdy*, two other gene functions map to *Sxr* and consequently to

the short arm: a gene involved in spermatogenesis termed *Spy* (ref. 26), and the gene controlling expression of the H-Y minor histocompatibility antigen *Hya* (ref. 27). *Sxr'* is a variant of *Sxr* that retains *Tdy* (XX *Sxr'* animals are male), but is deleted for *Hya* and *Spy*^{28,29}. *Sxr'* is therefore the minimum portion of the mouse Y chromosome known to contain *Tdy*.

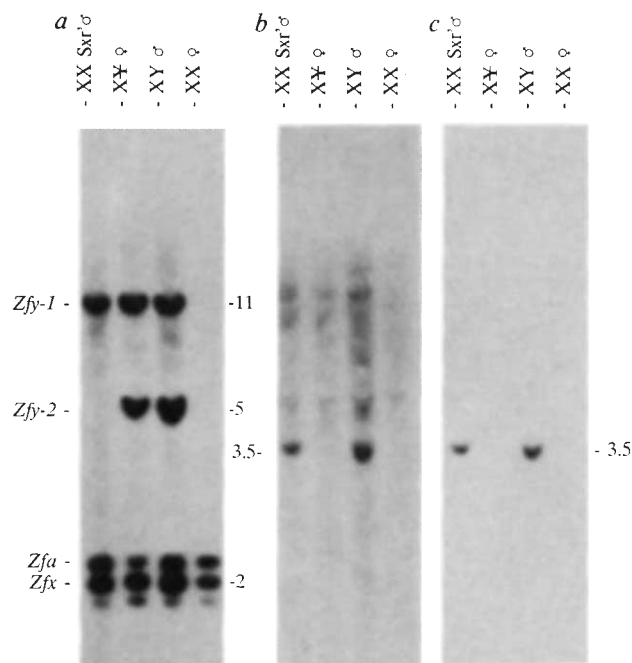


FIG. 1 Mapping conserved sequences to the sex-determining region of the mouse Y chromosome. A Southern blot of *EcoRI*-digested genomic DNA from an XX *Sxr'* male (a mouse carrying the minimum portion of the Y chromosome that confers maleness), an XY female (a mouse carrying a mutation in *Tdy*, ref. 35) and a normal XY male and XX female mouse, was hybridized consecutively to three separate probes. **a**, A 1.9-kb *HindIII* genomic fragment containing the zinc-finger domain of *Zfy-1* (ref. 32), which reveals bands corresponding to each of the four *Zfy*-related genes in XY male DNA. All of these genes are also seen in the XY female track. *Zfy-1*, but not *Zfy-2*, maps to *Sxr'* in the XX *Sxr'* track, but only the X-linked and autosomal members of this gene family (*Zfa* and *Zfx*) are present in a normal XX female. **b**, The sequence unique to the human sex-determining region, pY53.3¹⁹, which hybridizes strongly to a 3.5-kb band present in DNA from XY male and XX *Sxr'* male mice but absent from DNA from XX female and XY female mice. Additional weakly hybridizing bands are present in all tracks, so cannot be Y-linked. **c**, A mouse Y chromosome-derived clone, p422.04, containing sequences homologous to pY53.3, which hybridizes at high stringency only to the 3.5-kb Y-linked band. Fragment sizes are indicated (kb). **METHODS.** XY female mice carry a strain 129 Y chromosome, the same strain as the XX and XY samples used here. *Sxr'* was maintained on a C57B6J background. Mouse clone 4.2.2 was isolated from a size-selected library constructed from *EcoRI*-digested strain 129 male spleen DNA. DNA in the size range 3.0–4.0 kb was recovered from an agarose gel on GeneClean (Bio 101, Inc.), ligated into λ ZapII (Stratagene), and packaged with Gigapack gold packaging extract (Stratagene) using manufacturer's protocols. Plaque-forming units 0.5×10^6 were directly screened without amplification, using a 0.9-kb *HincII* fragment from pY53.3 as a probe and following previously described procedures⁴⁴. Positively hybridizing λ -clone 4.2.2 was plaque purified, and the insert recovered in pBluescript by *in vivo* excision from the λ ZapII vector as described by the manufacturer. A 380-bp *BglIII*-*PstI* fragment was isolated from p4.2.2 and subcloned into pBluescript (p422.04). This fragment contained all the homology to pY53.3 and was not repetitive. For Southern blots, restricted DNA was electrophoresed through 0.7% agarose and transferred to Hybond N (Amersham) in 0.5 M NaOH, 1.5 M NaCl. Probes were labelled with ³²P, added to the filter in $3 \times$ SSC buffer, 0.1% sodium pyrophosphate, 5 $\times$ Denhardt's solution, 0.1% sodium dodecyl sulphate (SDS), 9% dextran sulphate and 50 μ g ml⁻¹ denatured herring-testes DNA, and hybridized for 16 h at 65°C. The filter was washed in $0.1 \times$ SSC, 0.1% SDS for mouse-specific probes, and $2 \times$ SSC, 0.1% SDS for the human probe at 65°C, and autoradiographed for 1–3 days, then stripped in 0.1% SDS, 100°C between experiments.

A 2.1-kb *HindIII* DNA fragment termed pY53.3, that maps to within 8 kb of the human pseudoautosomal boundary, was identified in a human chromosome walk and found to contain sequences conserved on the Y chromosome of a range of mammalian species¹⁹. When pY53.3, or a 0.9-kb *HincII* fragment derived from it, was used as a probe on Southern blots of *EcoRI*-digested DNA from XY male and XX female mice, a strongly hybridizing 3.5-kb male-specific band was detected, in addition to several weaker bands present in both sexes. The same 3.5-kb band was also present in DNA from XX *Sxr'* males (Fig. 1b). Few sequences are known that map to the *Sxr'* region of the Y chromosome. Apart from *Zfy-1* (refs 30, 31) (see Fig. 1a) that has a so far undefined role in male germ-cell development^{32,33}, only repetitive sequences, such as *Bkm* (ref. 34) and *Sx1* (ref. 25) map to this region. The probe pY53.3 therefore defines a new single-copy sequence in the minimum fragment of the Y chromosome shown to be sex-determining in both mice and humans.

Further data were obtained from a line of sex-reversed female mice carrying a mutant Y chromosome. We have described a heritable mutation that gives rise to such XY female mice³⁵. The mutation segregates exclusively with the Y chromosome amongst the offspring of these females. Moreover, because of frequent non-disjunction of the X and Y chromosomes during female meiosis, the XY females often produce XXY daughters and XYY sons. The latter shows that the mutation can be

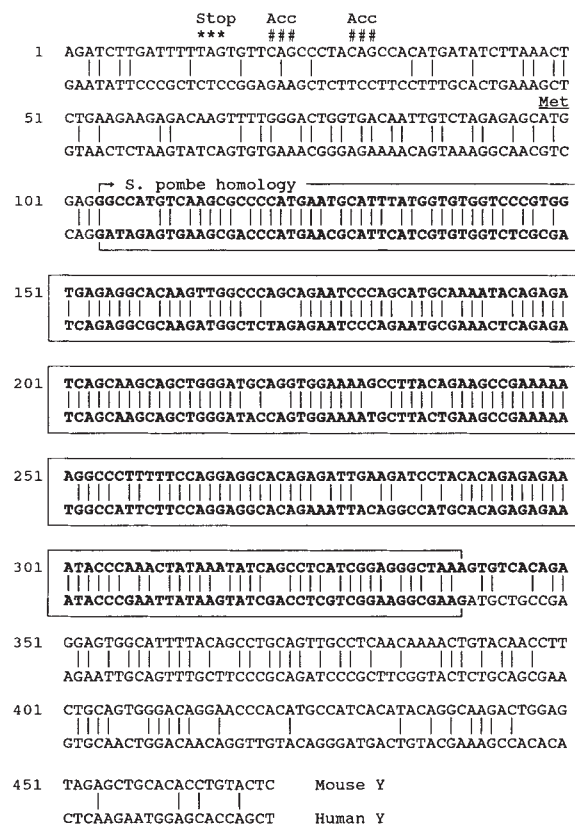


FIG. 2 Nucleotide comparison of mouse and human Y-linked sequences. The sequences of p4.2.2 (Mouse Y) and pY53.3 (Human Y) are shown on the upper and lower lines, respectively. Nucleotide homology is indicated by vertical bars. The single open reading frame in the mouse sequence is defined at the 5' end by a stop codon at nucleotides 14–16 (***). Within this open reading frame are two possible splice acceptor sites (##) and an in-frame translational start codon (Met). The region of amino-acid homology with the *S. pombe* Mc gene is boxed. Numbers, nucleotide positions.

complemented by a normal Y chromosome. Further characterization by karyotypic analysis and Southern blotting with a range of Y-specific DNA probes, suggested that there had been no gross deletion or rearrangement of the Y chromosome carrying the mutation. There also seemed to be no loss of Y-specific gene functions apart from that of testis determination, as the mutation was able to be fully complemented by *Sxr'*. Because of the phenotype and deduced location of the mutation, we concluded that it had occurred in *Tdy* itself. The mutation is therefore referred to as *Tdy^{m1}* and the Y chromosome carrying this mutation as *Y^{Tdy^{m1}}*, but for simplicity we have given this Y chromosome the symbol Υ .

We can therefore make a strong prediction that there must be a molecular basis for the *Tdy^{m1}* mutation in any candidate sequence for *Tdy*. *Zfy-1*, and its homologue *Zfy-2*, failed to satisfy this prediction as both genes have a normal structure and show a normal pattern of expression from the mutant Υ ³⁶. These genes are present in XY female mice (Fig. 1a). When the same filter is probed with pY53.3, the 3.5-kb *EcoRI* male-specific band is absent (Fig. 1b). Because there are no additional bands, this result shows that the hybridizing sequences have been deleted in the XY females. The probe pY53.3 is the first probe capable of distinguishing the mutant Υ chromosome from the normal Y chromosome, and at least has to be considered as the closest marker to *Tdy*. Although this does not constitute proof that the sequences detected by pY53.3 are part of the testis-determining gene, it is consistent with this possibility.

Cloning the mouse homologue of pY53.3

Several mouse genomic phage and cosmid libraries were screened with pY53.3. Although several clones were isolated,

none of them contained the 3.5-kb *EcoRI* fragment or mapped to the Y chromosome. They seemed instead to correspond to some of the fainter bands seen on Southern blots in both male and female DNA (see below). The difficulty in cloning the Y sequence may have been due to a general under-representation of Y-chromosomal DNA in genomic libraries or to a specific instability of the clone of interest. We therefore used size-selected DNA in a forced cloning strategy to isolate the 3.5-kb *EcoRI* Y-specific fragment (see legend to Fig. 1). Several clones of the expected size were obtained that hybridized strongly to pY53.3. One of these clones, p4.2.2, was analysed further and shown to contain a 380-base pair (bp) *BglII-PstI* restriction fragment (clone p422.04) that retained homology with pY53.3. When this 380-bp fragment was used as a probe at high stringency on genomic Southern blots, it hybridized only to the Y-specific 3.5-kb *EcoRI* band (Fig. 1c). Clone p4.2.2 therefore contains the mouse homologue of pY53.3.

Clone p4.2.2 was sequenced in the region of homology with pY53.3. Figure 2 shows a nucleotide comparison of the mouse and human sequences over 471-bp, in which the degree of homology is 62%. Of the six possible translation frames in p4.2.2, only one contained a long open reading frame in this region. This corresponds to one of the two open reading frames in pY53.3, and to an open reading frame in the corresponding Y-linked sequence from rabbit DNA¹⁹. The high degree of evolutionary conservation between the open reading frame region of p4.2.2 and that of pY53.3 is a strong indication that it is part of a functional gene. We have termed this gene *Sry* (a gene from the sex-determining region of the Y-chromosome).

The open reading frame in p4.2.2 contains near its 5' end two potential intron splice acceptor sites, and one in-frame ATG codon in a reasonable context for translational initiation³⁷. Poor

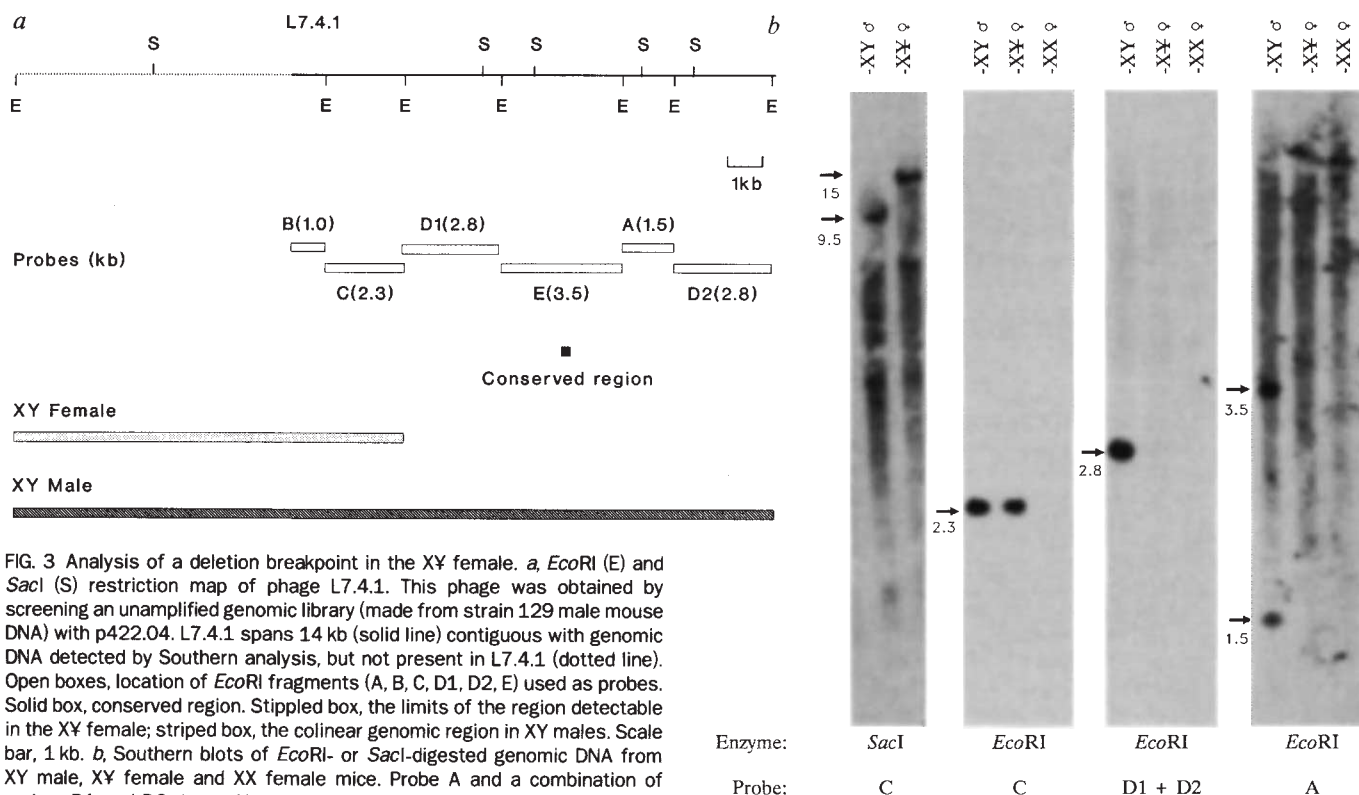


FIG. 3 Analysis of a deletion breakpoint in the XY female. *a*, *EcoRI* (E) and *SacI* (S) restriction map of phage L7.4.1. This phage was obtained by screening an unamplified genomic library (made from strain 129 male mouse DNA) with p422.04. L7.4.1 spans 14 kb (solid line) contiguous with genomic DNA detected by Southern analysis, but not present in L7.4.1 (dotted line). Open boxes, location of *EcoRI* fragments (A, B, C, D1, D2, E) used as probes. Solid box, conserved region. Stippled box, the limits of the region detectable in the XY female; striped box, the colinear genomic region in XY males. Scale bar, 1 kb. *b*, Southern blots of *EcoRI*- or *SacI*-digested genomic DNA from XY male, XY female and XX female mice. Probe A and a combination of probes D1 and D2 detect Y-specific bands in an XY male that are deleted in an XY female. Probe C detects a Y-specific *EcoRI* band of the same size in an XY male and an XY female. But a *SacI* digest reveals a difference in the size of the band detected by probe C, indicating a breakpoint within this genomic region in the XY female. Probe B, although found to be highly repetitive when used to probe a Southern blot of *SacI*-digested DNA, nevertheless gave the same result as probe C (data not shown). Sizes of the relevant bands (kb) are indicated.

METHODS. The library was constructed by partial digestion of genomic DNA with *Sau3a* and ligation into a λ FIX II vector (Stratagene) following manufacturer's protocols. The library was plated using DL652, a bacterial strain that stabilizes end to end repeats, to overcome our previous problems in cloning the mouse Y homologue, and screened as described⁴⁴. Southern blotting protocols as described in Fig. 1.

homology between the mouse and human sequences 5' to the ATG codon is consistent with functional constraints being lost at this point, although no corresponding ATG appears in the human sequence. Similarly, the potential splice acceptor sites in the p4.2.2 sequence have no cognates in pY53.3 (Fig. 2). There is evidence for the homologous region being the final exon of the human gene¹⁹, and we assume this also to be the case in *Sry*. A gradual divergence of the mouse and human sequences is seen towards the 3' end of the region of homology, implying a diminished functional importance of this part of the gene. No possible splice donor motifs have been detected at this end, nor has the 3' limit of the open reading frame been defined.

The most striking feature of the open reading frame in p4.2.2 is a 237-bp sequence in which the degree of homology with pY53.3 rises to 80% (Fig. 2). The conceptual translation of this sequence shows a similarity to the C-terminal 80 amino-acid residues of the Mc protein²¹. This homology is remarkable, given the evolutionary divergence between yeast and mammals, and indicates that these 237 bp define an important functional domain.

A deletion breakpoint in XY females

As the DNA fragment hybridizing to pY53.3, and the mouse homologue p4.2.2 are absent in XY females, it was necessary to clone a larger genomic region from the mouse to search for the boundaries of this deletion. Clone p422.04 was used to screen a male genomic library. From 10⁶ clones, one was obtained that showed very strong homology with p422.04. Restriction mapping confirmed that this clone, L7.4.1, contained the 3.5-kb *Eco*RI band corresponding to 4.2.2 and to the mouse Y-linked locus (data not shown). Figure 3a shows a restriction map of the insert from phage L7.4.1. Probes were made from each of the

*Eco*RI fragments contained within L7.4.1, and used to screen genomic blots of DNA from XY male, XY female and XX female mice. Three of the six probes (D1, D2 and A) failed to detect a Y-specific band in DNA derived from XY females (Fig. 3b). Note that probe A detects a 3.5-kb band in normal XY male in addition to the expected 1.5-kb band. This seems to be due to a cross-hybridizing sequence shared between probes A and E. Probe E also failed to detect either the 3.5-kb or the 1.5-kb band in DNA derived from XY females. Unlike probes D1, D2, A and E, probe C detects an *Eco*RI band of roughly the same size in DNA from both XY males and XY females. The adjacent probe B, which corresponds to one end of the phage insert, similarly detects an *Eco*RI band of the same size in DNA from both XY males and XY females (data not shown). These results suggest that the breakpoint is close to the *Eco*RI site between probes C and D1. This was confirmed by hybridizing *Sac*I-digested DNA with probe C. This reveals a band of altered size in XY females compared with XY males (also shown in Fig. 3b).

Expression of *Sry* in testis differentiation

Tdy is thought to act around 11.5 d.p.c. to divert the supporting cell precursors in the urogenital ridge from the follicle cell pathway to the Sertoli cell pathway. We have therefore used a sensitive polymerase chain reaction (PCR)-based assay to look for the presence of transcripts from *Sry* in urogenital ridges isolated from male and female embryos (Fig. 4). Oligonucleotide primers specific to *Sry* sequences in the conserved exon were used to amplify reverse transcribed RNA from 11.5 day urogenital ridges, adult testis and adult male liver. An *Sry* band is clearly seen in both adult testis and in male urogenital ridge, but not in liver or female urogenital ridge. The bands obtained are not from contaminating genomic or plasmid DNA in the samples, as control lanes in which reverse transcriptase was omitted show no signal. These results confirm that *Sry* is indeed an expressed gene. Further, *Sry* is expressed in the urogenital ridge at a time consistent with its having a role in testis determination.

Isolation of homologous cDNAs

Several complementary DNA libraries were screened, initially with pY53.3 and more recently with p422.04, to find cDNAs corresponding to *Sry*. In particular, the screening of two libraries prepared from 11.5-d.p.c. male genital ridges, a stage at which the testis-determining gene is thought to be expressed, failed to reveal any hybridizing phages; however, both libraries are very small (A.M., unpublished data, and ref. 38). The screen was therefore extended to 14.5 d.p.c. and adult testis libraries and to an 8.5-d.p.c. whole-mouse embryo library³⁹. Although the testis libraries have also failed so far to yield any recombinants with homology to the probe, the 8.5-d.p.c. library produced several positive clones. Detailed restriction mapping indicated that these clones fall into four discrete classes, but their individual use as probes on genomic Southern blots of male and female DNA indicated that none of them was Y-linked. We provisionally refer to these cDNAs as corresponding to autosomal loci, and assign to them the names a1-a4, although the possibility that they are X-linked has not been excluded. In each of these cDNAs a region was defined that showed strong hybridization to both the mouse and human probes and this region was therefore sequenced.

Sequence comparison defines a new 80-amino-acid motif with a high ratio of basic residues (up to 25%) and a strong helical content (Fig. 5). Conservation is strong in the motif, with 42 amino acids identical between the *Sry*-related genes. The amino-acid comparison reveals homology of this motif not only to the *S. pombe* Mc protein, but also to a domain present in four copies in hUBF (ref. 20). This domain, termed the HMG box because of its homology to the high mobility group proteins HMG1 and HMG2 (ref. 40), has been associated with binding to the

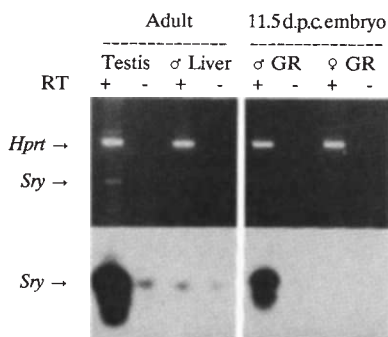


FIG. 4 PCR analysis of *Sry* expression in adult and embryonic mouse tissues. RNA was added to a reverse transcription reaction in the presence (+) or absence (-) of reverse transcriptase (RT). Subsequent PCR used *Sry*-specific oligonucleotide primers and included hypoxanthine phosphoribosyl transferase (*Hprt*) primers as a control for the quality and quantity of RNA in each sample. The upper panel shows an ethidium bromide-stained agarose gel with 148-bp PCR products corresponding to *Sry* transcripts in adult testis and, less intensely, genital ridge (GR) of 11.5-d.p.c. male embryos. The band was absent from adult male liver and 11.5-d.p.c. female genital ridge samples. *Sry* primers were derived from a single exon and are therefore capable of amplifying sequences from XY genomic DNA as well as from reverse transcribed *Sry* mRNA; the absence of visible product in the -RT samples confirms that any bands seen were due to the presence of *Sry* mRNA. *Hprt* primers spanning several exons do not amplify sequences from genomic DNA under the conditions used³², and yielded the expected 352-bp product in all +RT samples. The lower panel shows a corresponding Southern blot hybridized with probe p422.04. This confirms the *Sry* expression described above. In addition, a low level of genomic DNA contamination was revealed in some RNA samples; this does not affect the interpretation of the results. Independently isolated RNA samples have given similar results. METHODS. Total RNA (1 µg) was reverse transcribed and analysed by PCR as previously described³², except using 30 cycles and an annealing temperature of 53 °C. *Sry* primers were (5'-3') CTGTGTAAGATCTTCAATC and GTGGTGAGAGGCACAAGT.

upstream control element of the ribosomal RNA gene promoter to activate transcription²⁰. Clearly the motif present in the *Sry*-related genes is similar to, but distinct from, the HMG-box, thus defining a new family of genes. We have isolated five members of this family in the mouse (*SRY* and a1 to a4) but the number of bands seen when probing a Southern blot at low stringency with either pY53.3 or any of the cDNAs (data not shown) suggests that not all of the family has been cloned.

Certain amino-acid residues are identical throughout the *Sry*-related gene family and the Mc and HMG sequences (Fig. 5). This may be due to protein structural or sequence recognition constraints common to all genes of this type. Within the mouse gene family, at least three subfamilies can be distinguished (Fig. 5b). The sequences a1, a2 and a3 are almost identical in the conserved motif; the *Sry* and a4 sequences seem to have diverged independently from this group. Several residues are common to only the Y-linked sequences of the human, mouse and rabbit. The Y-linked gene products may therefore be functionally distinct from the autosomal gene products.

Discussion

Any candidate sequence for the testis-determining gene has to satisfy several criteria. In this and the accompanying paper by Sinclair *et al.*¹⁹, the isolation and sequence of part of a gene that would fulfil these criteria are described. First, it should be a sequence conserved on the Y chromosomes of all mammals shown to have a Y-chromosomal sex-determining mechanism. Southern blotting revealed a homologue of this sequence was present on the Y chromosome of a wide range of eutherian mammals. Second, it must map to the minimal portion of the Y chromosome shown to confer maleness. The human gene, termed *SRY*, was found in a detailed search of the 35 kb adjacent to the pseudoautosomal boundary, which is sufficient to confer maleness on an XX chromosomal background. We have shown here that the mouse homologue, *Sry*, maps to *Sxr'*, the smallest region of the Y chromosome known to contain *Tdy*. Third, a candidate must satisfy the prediction that there is a molecular basis for the occurrence of XY females shown genetically to

have a mutation in the testis-determining gene. We show here that part of *Sry*, including the highly conserved exon, is deleted in these *Tdy*^{m1} XY female mice. The finding of one deletion breakpoint close to the conserved exon is again consistent with *Sry* being *Tdy*, although this argument would clearly be strengthened if the other breakpoint were also found to be in or near the gene. In a previous study, we were unable to find any alteration in either the structure or expression of *Zfy-1* (ref. 36). This is the only other gene known to map to *Sxr'*, and had previously been considered a good candidate for *Tdy*.

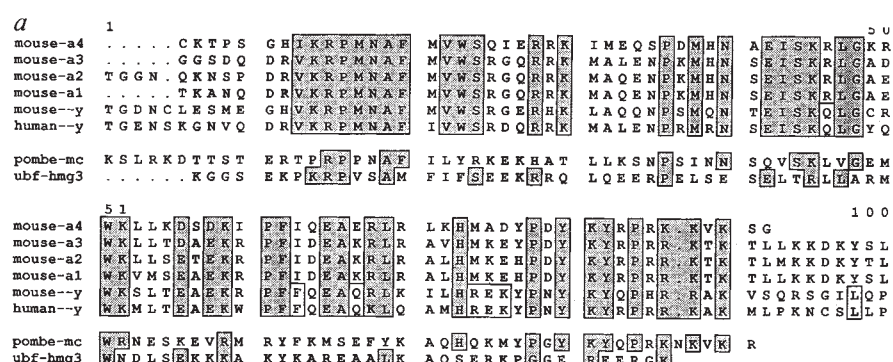
Finally, any candidate for *Tdy* should be expressed in a tissue and at a time consistent with what is known of its action. We can predict that *Tdy* should be expressed in the urogenital ridge at about 11.5-d.p.c., just before testis-cord formation, the first recognizable difference in male and female embryogenesis. The expression of *Sry* conforms with this prediction. The level of transcripts found in 11.5-d.p.c. urogenital ridge is very low, and preliminary PCR data also suggest that *Sry* is not expressed before 10.5 d.p.c. or during later stages of fetal gonad development (data not shown). This low level of expression is not surprising for a gene possibly operating as a switch in development.

Sry was also found to be expressed in adult testis by PCR (Fig. 4) and northern analysis (see accompanying paper¹⁹; also our results (data not shown)). Several genes suspected to have roles in developmental decisions in the embryo, for example, some homeobox-containing genes^{41,42}, also show expression in adult testis. *Sry* could have one role in the embryo in testis determination, and another postnatally in male germ-cell development.

If *Sry* proves to be involved in testis determination, it is not yet clear what its mode of action might be. Evidence suggests that *Tdy* promotes differentiation of Sertoli cells in a cell-autonomous manner⁴³. Further steps in male development follow directly from the differentiation of Sertoli cells. *Sry* might, therefore, be expected to encode a protein capable of acting as a regulatory molecule in the cell. The conserved protein domain in *Sry* is homologous to known DNA-binding proteins and, by

FIG. 5 Conservation of homologous protein domains. *a*, Comparison of amino-acid sequences (single-letter code) between homologous regions of the following sequences: ubf-hmg3, the third HMG box of hUBF²⁰, which shares the highest homology with this motif; pombe-mc, mating type protein Mc from *S. pombe*²¹; human-y, *SRY*; mouse-y, *Sry*; mouse-a1, mouse-a2, mouse-a3, mouse-a4, four different mouse cDNAs from genes not linked to the Y chromosome. Residues that are absolutely or strongly conserved between the upper six sequences are enclosed by shaded boxes. Where this conservation extends to the *S. pombe* Mc or hUBF-HMG proteins, these residues are also shaded. Residues that seem to be Y-specific in mouse, man (and rabbit¹⁹, not shown) are in open boxes. Numbers, amino-acid positions. The region of conservation defining a new protein motif extends from residue 11 to residue 90 in the upper sequences. *b*, The percentage of amino-acid homology, within the conserved motif, between the different sequences is summarized. The open box emphasizes the homology between the *Sry*-related sequences and both the *S. pombe* Mc sequence and the HMG motif, using the third HMG box of hUBF as an example. Shaded boxes emphasize the high homology between the mouse a-1, a-2 and a-3 sequences, as well as the lower degree of homology of the mouse a-4 sequence to the other members of the gene family.

METHODS. An 8.5-d.p.c. mouse total-embryo cDNA library³⁹ was screened with the *HincII* fragment of pY53.3. Fourteen positive clones were purified, and inserts subcloned into pBluescript (Stratagene). Restriction maps and Southern analysis of these clones showed that they correspond to four different loci not linked to the Y chromosome. For each of these messengers the region that hybridizes with pY53.3 was subcloned into pBluescript and sequenced as double stranded DNA using T7 polymerase (Pharmacia) according to the manufacturer's instructions. Sequence analysis and calculations



b

	Human Y	Mouse Y	Mouse a1	Mouse a2	Mouse a3	Mouse a4
UBF HMG3	47	47	53	53	50	49
Pombe Mc	54	56	58	58	58	56
Human Y		89	86	85	90	75
Mouse Y			81	80	82	77
Mouse a1				98	96	78
Mouse a2					95	78
Mouse a3						78

were performed using the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package⁴⁵.

analogy, the *Sry* protein could be a nuclear protein that binds DNA and acts as a transcriptional switch. The homology of *Sry* with the gene encoding the Mc protein at the *mat3-M* locus of *S. pombe* is intriguing. It is tempting to draw parallels between mating type in yeast and sex determination in mammals, although there is no reason to suppose any evolutionary connection between the mechanisms employed in the two cases. It is more likely that the conserved motif represents a functional protein domain that is appropriate for the type of control necessary to achieve the switch.

In attempts to isolate cDNAs corresponding to transcripts of *Sry*, we discovered at least four additional genes that contain the 80-amino-acid region of homology. The high degree of conservation of this protein box, both between the different

cDNAs and between the Y-linked genes in different species, implies that it is a functional domain. All four of the autosomal genes were isolated from an 8.5-d.p.c. embryo library, and could be involved in other early developmental decisions.

In conclusion, we have described a novel gene family in mice, linked by the presence of a conserved amino-acid domain also found in a gene involved in mating type of *S. pombe*, and in the DNA-binding protein hUBF. One member of this gene family maps to the sex-determining region of the Y chromosome and satisfies various predictions made for the testis-determining gene. We therefore consider this gene to be a good candidate for *Tdy*, although proof awaits functional tests of the gene in transgenic mice. □

Received 11 June; accepted 22 June 1990.

1. McLaren, A. *Trends Genet.* **4**, 153–157 (1988).
2. Ferguson, M. W. J. & Joenen, T. *Nature* **296**, 850–853 (1982).
3. Hodgkin, J. *Nature* **344**, 721–728 (1990).
4. Ford, C. E., Jones, K. W., Polani, P. E., de Almeida, J. C. & Briggs, J. H. *Lancet* **i**, 711–713 (1959).
5. Jacobs, P. A. & Strong, J. A. *Nature* **183**, 302–303 (1959).
6. Welshons, W. J. & Russell, L. B. *Proc. natn. Acad. Sci. U.S.A.* **45**, 560–566 (1959).
7. Burgoyne, P. S. *Phil. Trans. R. Soc. Lond.* **322**, 63–72 (1988).
8. McLaren, A. *Phil. Trans. R. Soc. Lond.* **322**, 3–9 (1988).
9. Borum, K. *Exp. Cell Res.* **24**, 495–507 (1961).
10. Burgoyne, P. S. *Nature* **342**, 860–862 (1989).
11. Merchants, H. *Dev. Biol.* **44**, 1–21 (1975).
12. McLaren, A., in *The Origin and Evolution of Sex* (eds Halvorsen, H. O. & Monroy, A.) 289–300 (Liss, New York, 1985).
13. Ferguson-Smith, M. A., Affara, N. A. & Magenis, R. E. *Development* **101** (Suppl.), 41–50 (1987).
14. Affara, N. A. *et al. Nucleic Acids Res.* **15**, 7325–7342 (1987).
15. Bishop, C. E., Boursot, P., Baron, B., Bonhomme, F. & Hatat, D. *Nature* **315**, 70–72 (1985).
16. Weissenbach, J., Leleu, J., Petit, C., Rouyer, F. & Simmler, M.-C. *Development* **101**, 67–74 (1987).
17. Palmer, M. S. *et al. Nature* **342**, 937–939 (1989).
18. Page, D. C. *et al. Cell* **51**, 1091–1104 (1987).
19. Sinclair, A. H. *et al. Nature* **346**, 240–244 (1990).
20. Jantzen, H.-M. *et al. Nature* **344**, 830–836 (1990).
21. Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. *EMBO J.* **7**, 1537–1547 (1988).
22. Cattaneach, B. M., Pollard, C. E. & Hawkes, S. G. *Cytogenetics* **10**, 318–337 (1971).
23. Evans, E. P., Burtenshaw, M. D. & Cattaneach, B. M. *Nature* **300**, 443–445 (1982).
24. McLaren, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6442–6445 (1988).
25. Roberts, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6646–6649 (1988).
26. Sutcliffe, M. J. & Burgoyne, P. S. *Development* **107**, 373–380 (1989).

27. Simpson, E., Edwards, P., Wachtel, S., McLaren, A. & Chandler, P. *Immunogenetics* **13**, 355–358 (1981).
28. McLaren, A., Simpson, E., Tomonari, K., Chandler, P. & Hogg, H. *Nature* **312**, 552–555 (1984).
29. Burgoyne, P. S., Levy, E. R. & McLaren, A. *Nature* **320**, 170–172 (1986).
30. Nagamine, C. M., Chan, K., Kozak, C. A. & Lau, Y.-F. *Science* **243**, 80–83 (1989).
31. Mardon, G. *et al. Science* **243**, 78–80 (1989).
32. Koopman, P. *et al. Nature* **342**, 940–942 (1989).
33. Nagamine, C. M., Chan, K., Hake, L. E. & Lau, Y.-F. *Genes Dev.* **4**, 63–74 (1990).
34. Singh, L., Phillips, C. & Jones, K. W. *Cell* **38**, 111–120 (1984).
35. Lovell-Badge, R. H. & Robertson, E. *Development* **109**, 635–646 (1990).
36. Gubbay, J., Koopman, P., Collignon, J., Burgoyne, P. & Lovell-Badge, R. *Development* **109**, 647–653 (1990).
37. Kozak, M. *Nucleic Acids Res.* **15**, 8125–8148 (1987).
38. Cunliffe, V., Koopman, P., McLaren, A. & Trowsdale, J. *EMBO J.* **9**, 197–205 (1990).
39. Fahrner, K., Hogan, B. L. M. & Flavell, R. *EMBO J.* **6**, 1265–1271 (1987).
40. Einck, L. & Bustin, M. *Exp. Cell Res.* **156**, 295–310 (1985).
41. Wolgemuth, D. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5813–5817 (1987).
42. Gaunt, S. J., Krumlauf, R. & Duboule, D. *Development* **107**, 131–141 (1989).
43. Burgoyne, P. S., Buehr, M., Koopman, P., Rossant, J. & McLaren, A. *Development* **102**, 443–450 (1988).
44. *Molecular Cloning: A Laboratory Manual* 2nd edn (eds Sambrook, J., Fritsch, E. F. & Maniatis, T.) (Cold Spring Harbor Laboratory, New York, 1989).
45. Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).

ACKNOWLEDGEMENTS. We thank P. Little for the bacterial strain DL652, and for advice on cloning difficult sequences. We also thank R. Braun and P. Rigby for helpful comments on the manuscript. A.M. is supported by the Deutscher Akademischer Austauschdienst and Dr Carl Duisberg Stiftung für das Auslandsstudium deutscher Studenten, B.C. by the NIH and J.C. by the Université Pierre et Marie Curie.

LETTERS TO NATURE

New observations of the cyclotron absorption feature in Hercules X-1

T. Mihara*, K. Makishima*, T. Ohashi*, T. Sakao*, M. Tashiro*, F. Nagase†, Y. Tanaka†, S. Kitamoto‡, S. Miyamoto‡, J. E. Deeter§ & P. E. Boynton||

* Department of Physics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

† Institute of Space and Astronautical Science, Sagami-hara, Kanagawa 229, Japan

‡ Department of Physics, Osaka University, Machikaneyama, Toyonaka, Osaka 560, Japan

§ Department of Astronomy and

|| Departments of Physics and Astronomy, University of Washington, Seattle, Washington 98195, USA

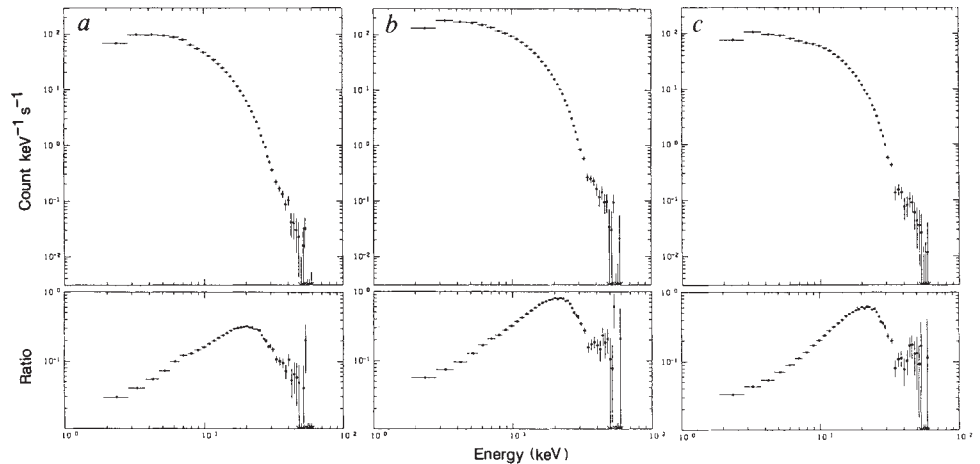
ALTHOUGH neutron stars are generally believed to be born with intense (10^{11} – 10^{13} G) magnetic fields^{1,2}, which then gradually decay³, measurements of their field strengths remain uncertain. In the special case of X-ray-emitting binary pulsars, a direct estimate of the field strength can be obtained by measuring the energy of spectral features that are due to electron cyclotron resonance^{4–13}. With the Ginga satellite observatory^{14,15}, we have measured a cyclotron feature in the hard X-ray spectrum of the 1.24-s binary pulsar Hercules X-1 with a much greater energy resolution than

in previous observations^{4–9}. The spectrum from 10–60 keV can be described with a simple analytical formula^{12,16,17}, which indicates an absorption feature at ~34 keV rather than an emission feature at ~50 keV. From this we estimate the surface magnetic field strength of Her X-1 to be $(2.9 \pm 0.3) \times 10^{12}$ G.

In mass-accreting pulsars, X-ray-emitting plasma is funnelled onto the magnetic poles^{1,2,11,13}, where the electron-photon interaction is dominated by the cyclotron resonance in the intense magnetic field $B \approx 10^{12}$ G. Theoretical work^{18–23} predicts that the resonance produces significant spectral features near the electron cyclotron energy, $E_a = 11.6(B/10^{12} \text{ G})$, which are more likely to appear in absorption than in emission. Such spectral features have been established in the two X-ray pulsars, Her X-1^{4–9} and 4U0115+63^{10,11}. From another X-ray pulsar, 4U1538-52^{12,13}, a cyclotron absorption feature was discovered at ~21 keV with the Large-Area Proportional Counter (LAC)¹⁵ on board Ginga. Despite the repeated observations of Her X-1^{4–9}, however, the spectral feature of this 'prototype' cyclotron object has not been established either as a ~35-keV absorption dip or as a ~50-keV emission line, although the former assignment is often favoured⁷. The LAC, with its wide spectral coverage and good energy resolution (8% at 35 keV), should be able to discriminate between these two.

We pointed the LAC at Her X-1 on 3 May, 3 June and 6 June 1989. An energy range up to 60 keV was attained by reducing the high voltage on the LAC to 1,760 V. The time resolution of the data was 500 ms and 62.5 ms in the May and June observations, respectively. Both in May and June, Her X-1 was found in the 'main on' state, the ~7-day X-ray-bright period

FIG. 1 X-ray pulse-height spectra of Her X-1 observed with the Ginga LAC. The detector response has not yet been removed. Lower panels show the spectra normalized to the spectrum of the Crab nebula (a power law of photon index 2.06) acquired with the same instrumentation. *a*, A spectrum averaged over the pulse phase, using both May and June data. Off-source background has been subtracted. *b*, Similar to *a*, but using only the pulse maxima from June observations. *c*, Difference between pulse maxima and minima for the June observations.



in its 35-d on-off cycle^{24,25}, with an intensity of ~ 90 mCrab. Figure 1a shows the X-ray pulse-height spectrum of Her X-1 using all the data, after background ('off-source') subtraction. The spectrum in Fig. 1b shows only the June data from the peaks of the 1.24-s pulses (13% duty ratio). The spectrum in Fig. 1c was derived from the June data as the difference between the pulse maxima and minima (62% duty ratio). In addition to the clear cutoff at ~ 20 keV, all the spectra exhibit noticeable structure at 30–40 keV which we identify with the cyclotron feature previously observed^{4–9}. The feature is more prominent in Fig. 1b than in Fig. 1a, justifying the way in which Fig. 1c was derived. This also ensures that the features are not the result of incorrect background subtraction.

To fit these wide-band spectra, we use a photon distribution of the form^{12,16,17} $f(E) = IE^{-\alpha} \exp(-H(E))$, where E is the photon energy, I is the normalization, and α is the photon index. The four-parameter function

$$H(E) = \frac{A_1(WE/E_a)^2}{(E - E_a)^2 + W^2} + \frac{A_2(2WE/2E_a)^2}{(E - 2E_a)^2 + (2W)^2} \quad (1)$$

simulates the cutoff and the cyclotron feature in terms of a common set of parameters. The first and the second terms of H represent the fundamental and the second-harmonic resonances, respectively; E_a is the energy of the fundamental resonance peak, A_1 and A_2 are the optical depths of the fundamental and the second-harmonic resonances, respectively, and W is the width of the feature. Inclusion of the second-harmonic term was

motivated by the detection of harmonic absorption lines from γ -ray bursts²⁶, and we arbitrarily assumed that the second-harmonic feature is twice as wide as the fundamental.

This form of H was derived for the classical optical depth owing to resonant cyclotron scattering in a uniformly magnetized cold plasma, and so may be too simple to represent the complicated radiative transfer taking place in the pulsar accretion column. We therefore do not claim that this model has physical significance, but it does give spectral shapes that are in qualita-

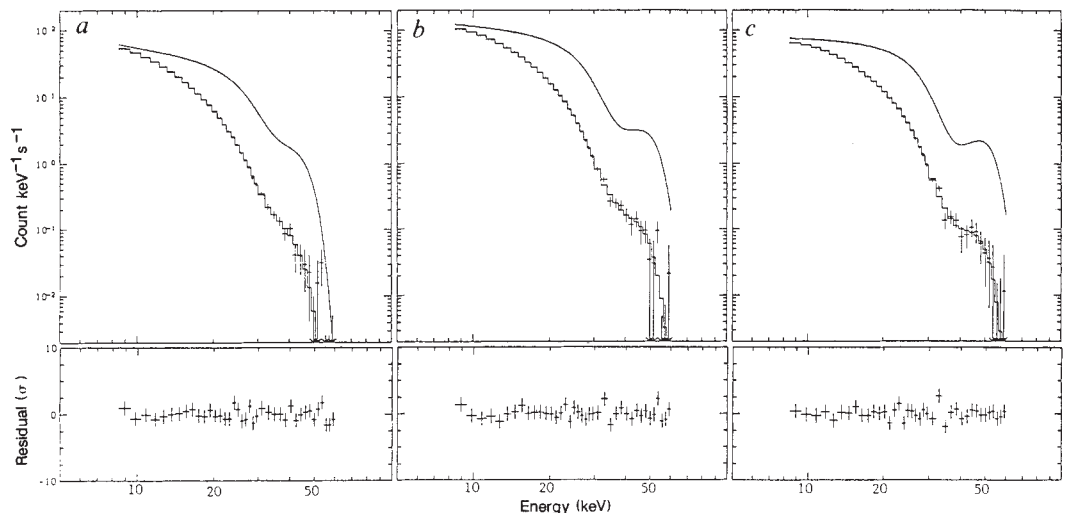
TABLE 1 Spectral parameters of Her X-1 in the 8.5–60 keV range

	Pulse phase average	Background* subtracted pulse	Pulse maxima minus pulse minima
Normalization /	206 ± 35	248 ± 41	92 ± 16
Photon index α	0.56 ± 0.07	0.32 ± 0.07	0.09 ± 0.07
Cyclotron energy E_a (keV)	32.5 ± 1.4	34.7 ± 0.9	35.4 ± 1.0
Width W (keV)	11.7 ± 2.7	12.0 ± 2.0	11.3 ± 2.0
Optical depth of fundamental A_1	1.7 ± 0.5	2.4 ± 0.3	2.8 ± 0.5
Optical depth of second harmonic A_2	13 ± 9	10 ± 6	12 ± 9
Reduced χ^2	0.81	0.85	0.73

Quoted errors are single-parameter 90% confidence limits. There are 33 degrees of freedom in each fit.

* 'Off-source' background.

FIG. 2 The 8.4–60 keV portion of the spectra of Her X-1 in Fig. 1, fitted with the cyclotron absorption model convolved with the detector response (histograms). The solid curves show the best-fit incident (corrected for the detector efficiency and resolution) spectra. The model agrees with the observed spectrum within the line thickness for energies below 20 keV. The fit residuals are displayed at the bottom of each panel.



tive agreement with the results of detailed numerical studies^{19–21}. Also it can describe the overall hard-X-ray spectrum of 4U1538–52 (ref. 12), including both the 21-keV absorption feature and the steep cutoff above ~15 keV. We therefore believe that the model can be used as an improved phenomenological formula to quantify the spectrum of X-ray-pulsars, and assume that the value of E_a obtained by fitting it to the data can be regarded as a reasonable estimate of the cyclotron resonance energy. We convolved spectra generated using this model with the detector response to compare with the observed spectra. The range of the fit was restricted to 8.5–60 keV, to avoid the complexities of the iron line and low-energy absorption¹³. Acceptable fits were obtained for all spectra, together with $E_a \approx 34$ keV (Fig. 2, Table 1).

The identification of the cyclotron feature of Her X-1 as an absorption feature at ~34 keV settles the long controversy about the nature of the feature, and yields a field estimate of $(2.9 \pm 0.3) \times 10^{12}$ G, although the value is somewhat model-dependent. (The mechanism, however, may be resonant photon scattering rather than pure absorption²³.) Also, the improved energy resolution of the LAC has allowed us to resolve the actual cyclotron profile (Fig. 2), which, in previous studies, was obtained indirectly and was therefore heavily model-dependent and uncertain. The single semi-empirical formula describing the cutoff and cyclotron absorption of Her X-1 can now be used as a convenient analytic representation in theoretical studies of radiation from X-ray-pulsars. □

Received 10 January; accepted 25 May 1990.

- Manchester, R. N. & Taylor, J. H. *Pulsars*, 231–235 (Freeman, San Francisco, 1977).
- Joss, P. C. & Rappaport, S. A. *A. Rev. Astr. Astrophys.* **22**, 537–592 (1984).
- Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. Not. R. astr. Soc.* **213**, 613–639 (1985).
- Trümper, J. *et al. Astrophys. J.* **219**, L105–L110 (1978).
- Voges, W. *et al. Astrophys. J.* **263**, 803–813 (1982).
- Coe, M. J., Engel, A. R., Quenby, J. J. & Dyer, C. S. *Nature* **268**, 508–509 (1977).
- Soong, Y., Gruber, D., Peterson, L. & Rothschild, R. *Astrophys. J.* **348**, 641–646 (1990).
- Tueller, J. *et al. Astrophys. J.* **279**, 177–183 (1984).
- Staubert, R. *et al. Space Sci. Rev.* **30**, 311–323 (1981).
- Wheaton, W. *et al. Nature* **282**, 240–243 (1979).
- White, N. E., Swank, J. H. & Holt, S. S. *Astrophys. J.* **270**, 711–734 (1983).
- Clark, G. W., Woo, J. W., Nagase, F., Makishima, K. & Sakao, T. *Astrophys. J.* **353**, 274–280 (1990).
- Nagase, F. *Publ. astr. Soc. Japan* **41**, 1–79 (1989).
- Makino, F. *et al. Astrophys. Lett. & Comm.* **25**, 223–233 (1987).

- Turner, M. J. L. *et al. Publ. astr. Soc. Japan* **41**, 345–372 (1989).
- Tanaka, Y. in *Radiation hydrodynamics in Stars and Compact Objects* (eds Mihalas, D. & Winkler, K. H.), 198–221 (Springer, Berlin, 1986).
- Makishima, K. *et al. Publ. astr. Soc. Japan* **42**, 295–315 (1990).
- Pravdo, S. H. *et al. Astrophys. J.* **225**, 988–993 (1978).
- Bonazzola, S., Heyvaerts, J. & Puget, J. L. *Astr. Astrophys.* **78**, 53–64 (1979).
- Mészáros, P. & Nagel, W. *Astrophys. J.* **298**, 147–160 (1985).
- Wang, J., Wasserman, I. & Salpeter, E. *Astrophys. J.* **338**, 343–358 (1989).
- Herold, H. *Phys. Rev. D* **19**, 2868–2875 (1979).
- Wang, J. C. L. *et al. Phys. Rev. Lett.* **63**, 1550–1553 (1989).
- Tananbaum, H. *et al. Astrophys. J.* **174**, L143–L149 (1972).
- Trümper, J., Kahabka, P., Ögelman, H., Pietsch, W. & Voges, W. *Astrophys. J.* **300**, L63–L67 (1986).
- Murakami, T. *et al. Nature* **335**, 234–235 (1988).

ACKNOWLEDGEMENTS. We thank F. Makino and other members of the Ginga team for their help. J.D. and P.B. were supported in part by NASA.

Measurement of the interactions between membranes in a stack

P. Richetti*, P. Kékicheff, J. L. Parker & B. W. Ninham

Department of Applied Mathematics, Australian National University,
GPO Box 4, Canberra ACT 2601, Australia

INTERACTIONS between membranes are important in many biological processes and are of increasing interest to biophysicists and physicists¹. The forces that determine the stability of a given configuration include van der Waals, electrostatic double-layer², hydration^{3–4}, undulation^{5,6}, ion-ion correlation⁷ and surface dipole-dipole⁸ forces, and their interplay depends critically on the system. Direct measurement of forces between two unsupported membranes in solution is a formidable task, but the interactions operating in a stack of membranes, such as a multilamellar mesophase, are not so inaccessible. The compressibility of the stack normal to the layers is related directly to the free energy density of interlayer interactions, so that measurements of the compressibility allow the type and contribution of the various forces to be inferred. Here we report the measurement of layer compressibility in a lamellar mesophase using a surface force apparatus⁹. The molecular resolution of this device also yields structural information on the confined mesophase, such as the reticular spacing and the nature and dynamics of packing defects.

Lytotropic lamellar mesophases are planar stacks of membranes that contain surfactant and/or cosurfactant and sometimes a solvent, separated by diluent. As early as 1941 Palmer *et al.*¹⁰ reported that brain lipids swell in water, but collapse on addition of salt. In other systems, the addition of diluent extends the reticular distance to micrometres^{11–13}, far beyond the range of electrostatic or solvation forces. Steric repulsion⁵ owing to the thermal fluctuations of the interfaces has been invoked to explain such swelling.

The techniques used to measure elastic properties of fluid lyotropic lamellar mesophases included high-resolution X-ray scattering^{14,15}, ellipsometry¹⁶, dynamic light scattering^{17,18},

response under mechanical constraint^{19,20} and electron paramagnetic resonance²¹. In particular, the modulus of layer compression B at constant chemical potential is related to the membrane interactions in the stack¹⁸ by

$$B = d \left(\frac{\partial^2 E}{\partial d^2} \right)_n \quad (1)$$

where E is the free energy per unit surface area, d the period, and n the number of layers per unit length ($n = 1/d$). Knowledge of $B(d)$ allows the type of interaction to be determined. For uncharged membranes or high concentrations of added electrolyte, undulation forces dominate and $B \propto 1/d^3$, but when electrostatic forces dominate $B \propto 1/d^2$ (ref. 22).

Previous techniques gave a combination of B and the layer-curvature elastic modulus^{14–16,19,21}, or results that were model-dependent^{17,18} or time-dependent owing to textural defects^{23,24}. Here these problems are avoided; B is derived directly from force measurements as a function of thickness for a multilamellar stack in a new type of surface force apparatus²⁵. The system we used for illustration was the lamellar mesophase formed by sodium dodecyl sulphate, pentanol and water (in proportions by weight²⁶ 7.17%, 17.48%, 75.35%; extra pure (BDH) and recrystallized from ethanol, distilled, and double distilled, respectively). X-ray¹⁵ and dynamic light scattering¹⁸ show that the interactions in this system are purely electrostatic. X-ray studies before and after the surface force measurements give a layer spacing d of 8.8 ± 0.1 nm for the sample.

The force across the lamellar phase confined between mica surfaces, with the layers mostly aligned parallel to these hard walls (homeotropic alignment), was measured 24 h after sample addition (Fig. 1). The forces oscillate irregularly with surface separations between 1,000 and 500 nm. As the separation is reduced further the periodicity becomes regular and close to the X-ray reticular distance, d . Only forces in regions where the slope of the force is less than the spring constant K can be measured⁹ and only part of each oscillation is accessible (see Fig. 1). The inward 'jump' from one stable position to the next between two branches occurred in a few seconds and we allowed the surfaces to stabilize at the new position for five minutes before continuing. The first 16 highly reproducible oscillations

* Permanent address: CRPP-CNRS, Château Brivazac, Avenue Schweitzer, 33600 Pessac, France.

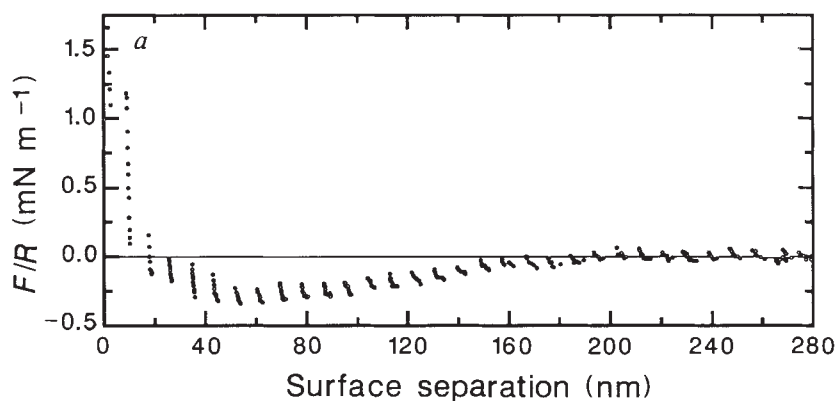
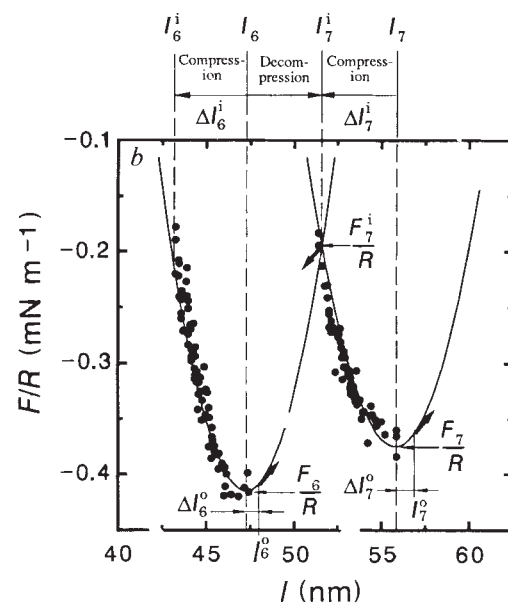


FIG. 1 *a*, Force F (normalized by the mean radius of curvature R of the surfaces) as a function of separation between two mica sheets immersed in a lamellar mesophase (volume 40 ml). The two partly silvered mica sheets are glued, silvered side down, onto the surface of two glass cylinders (radius R), aligned at right angles, one mounted on a piezoelectric tube, the other on a cantilever spring of stiffness K . The silvered surfaces form an optical cavity into which white light is directed and the surface separation is determined to ± 0.2 nm by the wavelengths emerging²⁸. The separation is adjusted mechanically or with the piezoelectric tube and determined to ± 0.1 nm. Note the occurrence of intrinsic unstable regimes without data when $\partial(F/R)/\partial l \geq K/R$, and the overall background attraction. *b*, Enlargement of the forces around the sixth and seventh oscillations (from contact) in both inwards and outwards runs. To account for the slight hysteresis in the voltage of the piezoelectric tube, polynomial fits were performed with the parts covered by both inward and outward runs and the remaining outward run values were extrapolated from them. The value of positions were not



affected. The separations were allowed to stabilize for several minutes after recording each jump in the outward run. The solid lines are parabolic fits to equation (2) using $B = 1.8 \times 10^6$ erg cm⁻³ and converting the measured forces to free energies under the Derjaguin approximation⁹. The large arrows show the theoretical jump positions. The inward jump positions l_n^i are observed to be located at the intersection of the two adjacent parabolas. The outward jump positions l_n^o are very close to the bottom of the parabola l_n . The theoretical difference $\Delta l_n^o = l_n^o - l_n$ is roughly 1% of the value of l_n ; for instance, $\Delta l_6^o \approx 0.5$ nm and $\Delta l_7^o \approx 0.6$ nm.

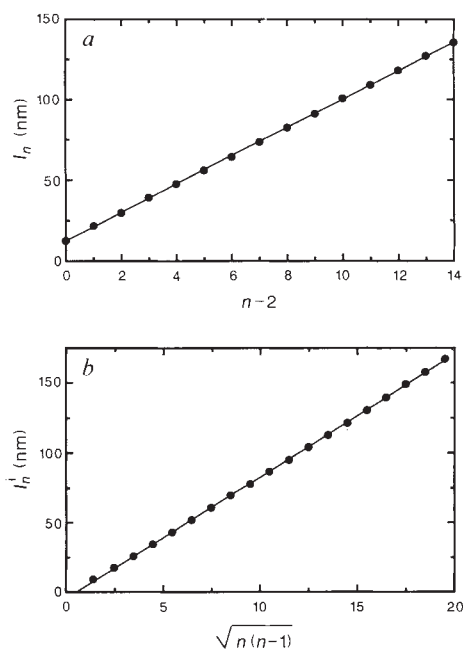


FIG. 2 For each oscillation n , the positions of the minima l_n are plotted against $n-2$ (*a*), and those of the inward jump l_n^i are plotted against $\sqrt{n(n-1)}$ (*b*). The solid lines are the least-square fits to the data giving a reticular distance of 8.8 nm in agreement with the X-ray measurement ($l_n = (8.8(n-2) + 12.4)$ nm and $l_n^i = (8.8\sqrt{n(n-1)} - 4.9)$ nm). The fact that the measured values are so close to the fitted line indicates that the forces were measured at the thermodynamic equilibrium. The first two points in the second series (*b*) lie slightly away from the line. This is because $d \neq d' = l_2/2 = 6.2$ nm and because the value of the compression modulus B' for $n=2$ (one bilayer confined between the surfaces) is different from the values obtained at larger n . Although the attractive background leads to free energies that are slightly different at the minima of each oscillation, this actually has a negligible effect on the l_n^i locations.

were tracked in both inward and outward runs using a piezoelectric device to position the mica surfaces (Fig. 1*b*). Between 20 and 200 nm the oscillations are superimposed on a net attractive background. At closer separations the force becomes repulsive, as for another swollen lamellar mesophase²⁷. Analysis of the ordinary refractive index obtained from the fringes of equal chromatic order²⁸ at the first three oscillations ($n = 1.33 \pm 0.02$, 1.350 ± 0.013 , 1.355 ± 0.011 at 2.8, 10.7, 18.5 nm, respectively) shows that no bilayer is present inside the first oscillation, a single bilayer exists between the surfaces in the second oscillation, two bilayers in the third, and so on.

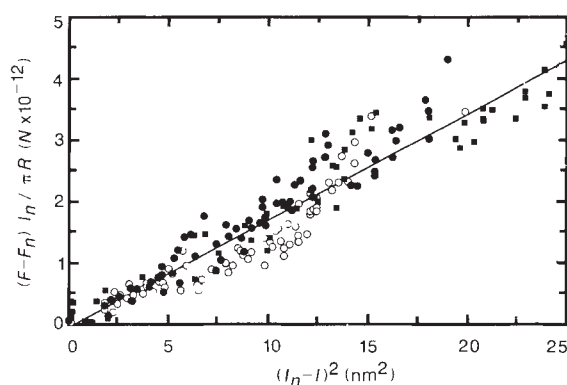
The stacked membranes are equivalent to a series of springs—each spring has the same stiffness, which depends on the adjacent membrane interactions, and a length corresponding to the reticular spacing d . When the separation between the mica surfaces changes, the stack experiences an elastic stress induced by the weak normal strain. For a layered system of thickness l_n at zero stress (an integral multiple of d), the corresponding elastic energy density has a parabolic shape centred around l_n

$$E = \frac{1}{2} B \frac{\Delta l_n^2}{l_n} \quad (2)$$

where $\Delta l_n = l_n - l$. The free energy density comprises a series of intersecting parabolas corresponding to the series of discrete layers. Two adjacent parabolas l_n^i intersect as

$$l_n^i = \sqrt{n(n-1)} d \quad \Delta l_n^i = (n - \sqrt{n(n-1)}) d \quad (3)$$

When $|\Delta l_n| > \Delta l_n^i$ the elastic energy of the system is lowered by the exclusion or introduction of a layer according to the sign. In other words the stress is released by the creation or annihilation of an elementary dislocation (a Burgers vector of 1). The analogous picture is the addition or removal of a spring in such a way that all the other components release their distortions to retain the net length. At large separations ($n \gg 1$) l_n^i is simply halfway between the two minima of adjacent parabolas $l_n^i \approx (n - \frac{1}{2})d$ and the distance that system can be distorted is $|\Delta l_n^i| \approx \frac{1}{2}d$ (ref. 29).



In the apparatus, reduction of the surface separation induces an elastic stress in the liquid crystal, that is, the system undergoes a compression, the solvent is expelled, and the left side of the parabola is followed, as shown in Fig. 1b. The solid parabolic lines are obtained from equation (2) with $B = 1.8 \times 10^6 \text{ erg cm}^{-3}$. Here the Derjaguin approximation⁹, which relates the interaction free energy E between flat surfaces of unit area to the force F normalized by the mean curvature R of the surfaces ($E = F/2\pi R$), is assumed (the validity of this approximation for multilayered systems³⁰ will be established elsewhere). The parabola is followed until intersection occurs with the next parabola (Fig. 1b). At this point the system would follow the decompression part of the next parabola (a layer being squeezed out by the creation of an edge dislocation), but an inward jump to the compressive part of the next oscillation occurs because of the instability of the cantilever spring ($\partial(F/R)/\partial l \geq K/R$). When the movement of the surfaces is reversed the force is measured past the minimum of the n th parabola and the surfaces jump over the decompression part when the slope of the force exceeds K (inclusion of a supplementary layer by annihilation of a dislocation).

The distance d' between the first bilayer and the mica surface (observed $2d' = 12.4 \text{ nm}$) is slightly less than the reticular spacing d so the positions of the minima l_n are a linear function of $n-2$: $l_n = 2d' + (n-2)d$ (Fig. 2a). The inward jump positions l_n^i cannot exceed $\sqrt{n(n-1)}d$ and, if the system is at equilibrium, should be a linear function of $\sqrt{n(n-1)}$, as is clear from Fig. 2b. The circles denote experimental values, and the solid lines are the best linear fits to these values. The slopes, 8.8 nm , agree with the reticular spacing obtained from X-ray diffraction.

Using equation (2) and the Derjaguin approximation⁹ the compressibility modulus B and measured force are related by

$$(F - F_n) \frac{l_n}{\pi R} = B \Delta l_n^2 \quad (4)$$

Here the free energies $F_n/2\pi R$ at the minima l_n , are non-zero because of an attractive background. We checked that B is not dependent on n by plotting equation (4) for $n = 5, 6, 12$ (Fig. 3). These points lie about a straight line of slope $B = 1.7 \times 10^6 \text{ erg cm}^{-3}$. A more refined analysis, taking into account both the total energy expended on going from l_n to l_n^i for the first 16 oscillations and the decrease in d for the two first layers gives $B = (1.9 \pm 0.4) \times 10^6 \text{ erg cm}^{-3}$, in close agreement with that inferred from high-resolution X-ray scattering¹⁵.

Our experimental value also agrees with the theoretical one. The electrostatic pressure exerted on a membrane is known². Using equation (1) for a system with a high density of charges and large reticular distance, we have¹⁵

$$B = \frac{\pi^2 k_B T d}{2L(d-\delta)^3} \left[1 - 3 \frac{\Sigma}{\alpha L(d-\delta)} + 6 \frac{\Sigma^2}{\alpha^2 L^2 (d-\delta)^2} + \dots \right] \quad (5)$$

where δ is the bilayer thickness (2.0 nm , ref. 15), α is the dissociation coefficient (assumed to be unity), Σ is the surface area per charged polar head ($\Sigma \approx 2 \text{ nm}^2$, deduced from the

FIG. 3 The data from the fifth (●), sixth (○), and twelfth (□) oscillations plotted according to equation (4). For $n=2$, the compressional modulus B' departs from the bulk value B . It can be estimated using equation (5) for an equivalent system with $d' = 6.2 \text{ nm}$ and the same values for δ and Σ (constant Σ is not strictly valid but is a reasonable approximation) using these parameters $B' = 2.68 \times 10^6 \text{ erg cm}^{-3}$. The system with $n \geq 2$ layers has an effective $B_n = l_n B B' / (2d'B + (n-2)dB') = \gamma(n)B$ which is a combination of B and B' . $B' \approx 3/2B$, so the correction $\gamma(n) \approx (3l_n)/(3(n-2)d + 4d')$ seems to be $<10\%$ (less than the experimental error) for $n > 5$. A linear fit to $(F - F_n)/R = B\gamma(n)\pi\Delta l_n^2/l_n$ for the data of the first 16 oscillations gives $B = (1.9 \pm 0.4) \times 10^6 \text{ erg cm}^{-3}$.

composition of the system), and $L = \pi e^2 / \epsilon k_B T$ is a characteristic length ($L = 2.2 \text{ nm}$ at room temperature). With these parameters we obtain $B = 1.82 \times 10^6 \text{ erg cm}^{-3}$, in remarkable agreement with the measured value.

The attractive background has only a very small effect on the measurement of B because its variation is slow compared with the oscillation amplitudes. The existence of background attraction in all the systems studied so far²⁷ is surprising. An explanation, invoking disjoining pressure has been proposed²⁷, but another is possible: in the crossed-cylinder geometry, topologically equivalent to a sphere on a flat surface, there exists an array of circular edge dislocations. Between two adjacent dislocations, the layered system experiences compression in some parts of the cell and decompression at others because there is a fixed number of layers for a continuously varying cell thickness. Owing to the curved wall the net balance of the stresses might be favourable to expansion, leading to an effective attraction between the surfaces.

We have shown that the direct measurement of the layer compressibility modulus in smectic phases is feasible. It is now possible to define the interactions between bilayers and the dependence of B as a function of the reticular distance along a dilution pathway. At high dilution (large spacing d) the measurement of B might be more difficult (B scales as $1/d^2$ or $1/d^2$) because of the resolution of the apparatus. Elastic features in other structured phases or in thermotropic liquid crystals are also accessible, as are intramembrane forces reflected in bending moduli. □

Received 2 January; accepted 16 May 1990.

1. Rand, R. P. *A. Rev. Biophys. Bioeng.* **10**, 277-314 (1984).
2. Cowley, A. C., Fuller, N. L., Rand, R. P. & Parsegian, V. A. *Biochemistry* **17**, 3163-3168 (1978).
3. LeNeveu, D. M., Rand, R. P., Parsegian, V. A. & Gingell, D. *Biophys. J.* **18**, 209-230 (1977).
4. Pashley, R. M. & Israelachvili, J. N. *Colloids & Surfaces* **2**, 169-187 (1981).
5. Helfrich, W. *Z. Naturforsch. A33*, 305-315 (1978).
6. Peliti, L. & Leibler, S. *Phys. Rev. Lett.* **54**, 1690-1693 (1985).
7. Kjellander, R., Marčelja, S., Pashley, R. M. & Quirk, J. P. *J. Phys. Chem.* **92**, 6489-6492 (1988).
8. Attard, P., Mitchell, J. D. & Ninham, B. W. *Biophys. J.* **53**, 457-460 (1988).
9. Israelachvili, J. N. & Adams, G. E. *JCS Faraday Trans. 1* **74**, 975-1001 (1978).
10. Palmer, K. J. & Schmitt, F. O. *J. cell. comp. Physiol.* **17**, 385-394 (1941).
11. Larché, F., Appel, J., Porte, G., Bassereau, P. & Marignan, J. *Phys. Rev. Lett.* **56**, 1700-1703 (1986).
12. Lichterfeld, F., Schmeling, T. & Strey, R. *J. Phys. Chem.* **90**, 5762-5766 (1986).
13. Satoh, N. & Tsujii, K. *J. Phys. Chem.* **91**, 6629-6632 (1987).
14. Safinya, C. R. *et al. Phys. Rev. Lett.* **57**, 2718-2721 (1986).
15. Roux, D. & Safinya, C. R. *J. Phys., Paris* **49**, 307-318 (1988).
16. Meunier, J. *J. Phys., Paris* **46**, L1005-L1007 (1985).
17. Nallet, F., Roux, D. & Prost, J. *Phys. Rev. Lett.* **62**, 276-279 (1989).
18. Nallet, F., Roux, D. & Prost, J. *J. Phys., Paris* **50**, 3147-3156 (1989).
19. Nallet, F. & Prost, J. *Europhys. Lett.* **4**, 307-313 (1987).
20. Oswald, P. & Allain, M. *J. Phys., Paris* **46**, 831-838 (1985).
21. Di Mèglio, J. M., Dvornitzky, M. & Taupin, C. *J. Phys. Chem.* **89**, 871-874 (1985).
22. Liebler, S. & Lipowsky, R. *Phys. Rev. B35*, 7004-7009 (1987).
23. Pershan, P. S. & Prost, J. *J. appl. Phys.* **46**, 2343-2353 (1975).
24. Bartolino, R. & Durand, G. *Phys. Rev. Lett.* **39**, 1346-1349 (1977).
25. Parker, J. L., Christenson, H. K. & Ninham, B. W. *Rev. sci. Instrum.* **60**, 3135-3138 (1989).
26. Belloq, A. M. & Roux, D. in *Microemulsions* (eds Frieberg, P. & Botherel, P.) (Chemical Rubber Company, Boca Raton, 1987).
27. Kélicheff, P. & Christenson, H. K. *Phys. Rev. Lett.* **63**, 2823-2826 (1989).
28. Israelachvili, J. N. *J. Colloid & Interface Sci.* **44**, 259-272 (1973).
29. Pershan, P. S. *J. appl. Phys.* **45**, 1590-1604 (1974).
30. Horn, R. G., Israelachvili, J. N. & Perez, E. *J. Phys., Paris* **42**, 39-52 (1981).

ACKNOWLEDGEMENTS. We thank D. Roux, F. Nallet, P. Barois and J. N. Israelachvili for encouragement and discussions, and R. M. Pashley and H. K. Christenson for technical assistance and advice.

Complete reduction of carbon dioxide to carbon using cation-excess magnetite

Yutaka Tamaura & Masahiro Tabata

Department of Chemistry, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo 152, Japan

THE reduction of gaseous oxides such as CO_2 and H_2O is an important concern in industrial processes and pollution control. Here we report the reduction of carbon dioxide to carbon with an efficiency of nearly 100% at 290 °C using cation-excess magnetite ($\text{Fe}_{3+\delta}\text{O}_4$, $\delta = 0.127$). In this reaction, the oxygen in the CO_2 is transferred, in the form of O^{2-} , to the cation-excess magnetite, and no gas is evolved. The carbon in the CO_2 is reduced to carbon (zero valence) by the addition of an electron donated from the cation-excess magnetite to maintain electrical neutrality during the transfer of the O^{2-} to the magnetite. When we used H_2O in place of CO_2 , hydrogen gas was evolved, indicating that the same mechanism can also reduce H_2O .

The stoichiometric mixed-valence compound magnetite, Fe_3O_4 , gives an X-ray diffraction pattern characteristic of the spinel structure¹. The magnetite powder (100–200 nm) that we synthesized by air oxidation of an Fe(II) hydroxide suspension^{2,3} gave a similar X-ray diffraction pattern and contained no other peaks. The chemical composition of this powder, however, was $\text{Fe}_{2.887}\text{O}_4$. This cation deficiency is considered to be due to oxidation of some of the low-valency iron in particles <1 μm in size^{3–5}. The Mössbauer spectrum of our compound showed evidence of oxidation of some of the Fe(II) in the B sites of the spinel structure—the ratio of the peak areas corresponding to the octahedral (Fe(III)) and tetrahedral (Fe(II) + Fe(III)) sites was less than two, the value characteristic of stoichiometric magnetite^{4–6}. Also, careful oxidation of the small particles (<1 μm) of Fe_3O_4 yields cation-deficient magnetite ($\text{Fe}_{3+\delta}\text{O}_4$; $\delta < 0$)⁴. The lattice constant of the cation-deficient magnetite was 0.8389 nm (curve C in Fig. 1), characteristically lower⁵ than that of stoichiometric Fe_3O_4 (0.83967 nm)⁷.

Although much is known about the properties of cation-deficient magnetite⁸, the data for cation-excess magnetite are very limited^{8,9}. Dieckmann⁸ has studied cation-excess magnetite in the temperature range 900–1,400 °C, where it exists in equilibrium with wüstite. But there has been no report on cation-excess magnetite below 500 °C (where stoichiometric magnetite exists in equilibrium with $\alpha\text{-Fe}$ ^{10,11}), probably because conditions under which the cation-excess compound is stable have not yet been found.

We expected that the reaction of magnetite with hydrogen

would form cation-excess magnetite by releasing oxygen. Therefore we heat-treated (290 °C) 3 g of magnetite powder for 4 h, using a hydrogen flow of 200 ml min⁻¹ through a quartz tube reactor (300 mm long, 25-mm inner diameter). After the heat treatment, we quenched the sample using a cold bath of ice and NaCl. This hydrogen-treated magnetite was stable in a nitrogen atmosphere (Table 1). Its X-ray diffraction pattern, taken under conditions to protect it from oxidation, contains only the peaks of the spinel-type compound. The lattice constant was 0.8418 nm (curve A in Fig. 1), which is larger than that of the stoichiometric Fe_3O_4 (0.83967 nm)⁷ owing to the presence of excess iron in the interstices. In thermogravimetric (TG) analysis using a flow of hydrogen, weight loss was observed in the range 250–300 °C. Gas chromatographic analysis of the outlet gas stream showed that the weight loss was accompanied by the release of the H_2O . The chemical composition of the hydrogen-treated magnetite was determined to be $\text{Fe}_{3+\delta}\text{O}_4$ ($\delta = 0.127$) by chemical analysis. This composition is nearly equal to that estimated from the thermogravimetric analysis, assuming that the weight loss is due to the removal of the oxygen ion.

When the cation-excess magnetite was heated at 400 °C under a nitrogen atmosphere, X-ray diffraction peaks corresponding to $\alpha\text{-Fe}$ appeared along with those of the spinel-type compound, showing that the cation-excess compound is metastable at 400 °C. As mentioned above, however, it is stable under nitrogen at room temperature (Table 1).

In air at room temperature cation-excess magnetite gradually becomes oxidized and changes into cation-deficient magnetite (Table 1). X-ray diffractometry showed that the cation-deficient magnetite is composed of the iron oxides of the solid solution of Fe_3O_4 and $\gamma\text{-Fe}_2\text{O}_3$. Thus, metastable cation-excess magnetite seems to be able to incorporate oxygen in the form of O^{2-} into the oxygen sublattice of the magnetite. The incorporation of O^{2-} will generate a driving force to donate the electron and preserve the electrical neutrality of the magnetite, and we therefore expected that the cation-excess magnetite would be able to remove the oxygen from oxides, and reduce the oxidized element in them.

We tested this hypothesis using CO_2 . After generating the cation-excess magnetite as described above, we passed CO_2 gas through the reaction cell for 5 min. After confirming (by gas chromatography) that the H_2 in the cell had been completely

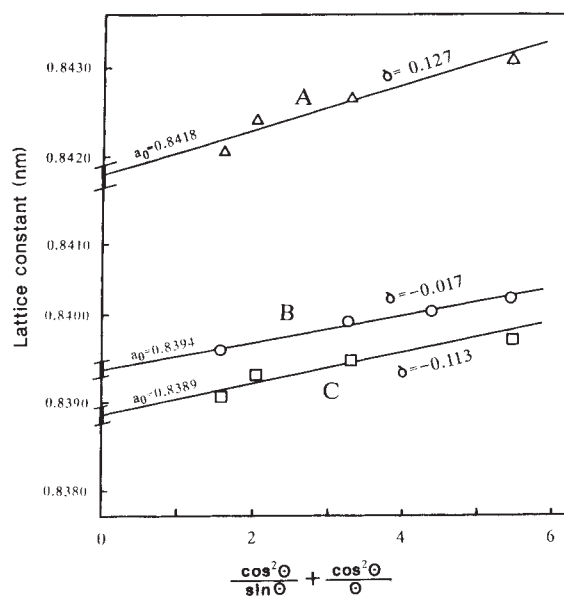


FIG. 1 The relationship between the lattice constant and the Nelson-Riley function for $\text{Fe}_{3+\delta}\text{O}_4$; after the H_2 treatment (curve A), after the reaction A with CO_2 (B), and before the H_2 treatment (C).

TABLE 1 Composition of magnetite in air and nitrogen

Time (min)	Chemical composition of $\text{Fe}_{3+\delta}\text{O}_4$ *
in air (25 °C)	
0	$\text{Fe}_{3.127}\text{O}_4$ ($\delta = 0.127$)
1	$\text{Fe}_{3.020}\text{O}_4$ ($\delta = 0.020$)
3	$\text{Fe}_{2.956}\text{O}_4$ ($\delta = -0.044$)
8	$\text{Fe}_{2.903}\text{O}_4$ ($\delta = -0.097$)
in N_2 (25 °C)	
5	$\text{Fe}_{3.127}\text{O}_4$ ($\delta = 0.127$)
100	$\text{Fe}_{3.127}\text{O}_4$ ($\delta = 0.127$)
600	$\text{Fe}_{3.127}\text{O}_4$ ($\delta = 0.127$)

* X-ray diffractometry showed that $\text{Fe}_{3+\delta}\text{O}_4$ was the cation-excess magnetite ($\delta > 0$) or the cation-deficient magnetite ($\delta < 0$). The δ -values were determined by the chemical analysis (2,2'-bipyridyl method¹⁵), which were nearly equal to those estimated from the weight change in the gravimetry.

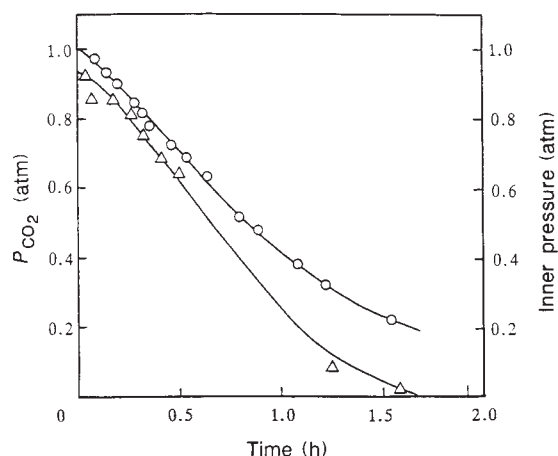


FIG. 2 The decrease in p_{CO_2} (triangles) and inner pressure (circles) as a function of time for the reaction between CO_2 and $\text{Fe}_{3+\delta}\text{O}_4$ ($\delta=0.127$).

replaced by CO_2 , we closed the inlet and outlet valves. (We define this as the zero time.) The inner pressure was measured with a pressure gauge, and the gas species analysed by gas chromatography. The decrease in CO_2 matches the decrease in the inner pressure (Fig. 2), indicating that no gas products are formed in the reaction and therefore that the CO_2 is decomposed into carbon. Gas chromatography showed that the only gaseous product was a small amount of CO . The carbon element analysis (Perkin Elmer 2400 CHN analyser) showed that the amount of carbon deposited on the magnetite was nearly equal to that present in the reaction cell. The X-ray diffraction pattern of the magnetite powder after the reaction, contained only peaks characteristic of the spinel-type compound. The lattice constant and the chemical composition of the magnetite after the reaction were 0.8394 nm (curve B in Fig. 1) and $\text{Fe}_{2.983}\text{O}_4$, respectively, nearly equal to those of the stoichiometric Fe_3O_4 ^{1,7}. The reaction rate (determined from the tangent to the curve of CO_2 partial pressure as a function of time) depended on the amount of the cation-excess magnetite, which may be due to a change in the surface area of the oxygen-deficient magnetite. At 290°C the rate was found to be $2.9\text{--}3.5 \times 10^{-3}\text{ mol per min per g cation-excess magnetite}$.

The decomposition rate is highest for the greatest oxygen deficiency, indicating that the transfer of oxygen from CO_2 to the magnetite is closely related to the mobility of the oxygen ion in the iron oxide. It is the oxygen transfer that gives rise to the high reduction potential of the Fe(II) ions for CO_2 . The broadening of the peaks in the X-ray diffraction pattern indicates that this high reduction potential is due to the instability of the spinel structure of the cation-deficient magnetite, which is distorted compared with the structure of the stoichiometric compound. When we used H_2O in place of CO_2 , hydrogen gas was evolved, indicating that the cation-excess magnetite would reduce H_2O by the same mechanism as that for CO_2 .

Copperthwaite *et al.*¹² have studied the decomposition reaction of CO_2 on the surface of the metals using X-ray photoelectron spectroscopy, but there have been no reports on the complete decomposition of CO_2 with an efficiency approaching 100%. We found that no CO_2 was decomposed below 250°C with the metals such as Mg , Al and Cu , and that only a small amount of CO_2 was decomposed at 290°C (3% after 6 h of reaction). In the Bosch reaction^{13,14}, in which CO_2 and H_2 are reacted with iron metal at $426\text{--}726^\circ\text{C}$, byproducts such as CO and CH_4 are formed, and the decomposition efficiency is $<10\%$ ¹⁴. In these reactions with the metals, the crystal structures of the metals change to those of the metal oxides. The high decomposition efficiency of the cation-excess magnetite is due to the fact that its spinel structure is kept during the reaction. □

Received 30 April; accepted 25 May 1990.

1. Smith, J. & Wijn, J. I. J. *Ferrite*, 136 (Philips Technical Library, Tokyo, 1965).
2. Kiyama, M. *Bull. chem. Soc. Japan* **47**, 1646–1650 (1974).
3. Katsura, T. & Tamura, Y. *Bull. chem. Soc. Japan* **52**, 96–100 (1979).
4. Annersten, H. & Hafner, S. S. Z. *Kristallogr.* **137**, 321–340 (1973).
5. Volenik, K., Seberial, M. & Neid, J. *Czech. J. Phys.* **B25**, 1063–1071 (1975).
6. Topsøe, H., Dumesic, J. A. & Boudart, M. *J. Phys., Paris* **C6**, 411–413 (1974).
7. JCPDS card 19-629 (Joint Committee on Powder Diffraction Standards, Swarthmore, 1989).
8. Diekmann, R. *Ber. Bunsenges. phys. Chem.* **86**, 112–118 (1982).
9. Diekmann, R. & Schmalzried, H. *Ber. Bunsenges. phys. Chem.* **81**, 414–419 (1977).
10. Darken, L. S. & Gurry, R. W. *J. Am. chem. Soc.* **67**, 1398–1412 (1945).
11. Darken, L. S. & Gurry, R. W. *J. Am. chem. Soc.* **68**, 798–816 (1945).
12. Copperthwaite, R. G., Davies, P. R., Morris, M. A., Roberts, M. W. & Ryder, R. A. *Catal. Lett.* **1**, 11–19 (1988).
13. Manning, M. P. & Reid, R. C. *ASME Pap.* 75-ENAS-22 (*Am. Soc. mech. Engineers*, 1975).
14. Wagner, R. C., Carrasquillo, R., Edwards, J. & Holmes, R. *18th Intersoc. Conf. Environmental Systems*, SAE Tech. Pap. Ser. 880995, 1–9 (Soc. Automotive Engineers, 1988).
15. Iwasaki, I. *et al. Bull. volc. Soc. Japan, Ser. II* **5**, 9–24 (1960).

Production of hydrogen peroxide in forest air by reaction of ozone with terpenes

Karl H. Becker*, Klaus J. Brockmann† & Joseph Bechara*

* Physikalische Chemie FB9, Bergische Universität, 5600 Wuppertal 1, FRG

† Fraunhofer Institut für Atmosphärische Umweltforschung, 8100 Garmisch-Patenkirchen, FRG

HYDROGEN peroxide and other hydroperoxides are thought to play a part in forest decline^{1–7}. The toxic effects of these compounds on plants are well documented^{1–4} and some hydroperoxides, including H_2O_2 , have been detected in forest air^{5,6} and in plant leaves⁷. Higher concentrations of H_2O_2 have been found inside forests compared with those measured close to the forest perimeter⁸. The reaction of atmospheric ozone with the isoprene and terpenes emitted by vegetation results in the formation of a wide variety of radicals^{9,10} whose subsequent reactions can contribute to the formation of hydroperoxides. We have used tunable-diode laser absorption spectroscopy to investigate the reaction of ozone with some alkenes, including isoprene and some monoterpenes, to see whether H_2O_2 is formed under a variety of atmospheric conditions. All the reactions studied produced H_2O_2 and a significant increase in the yields of H_2O_2 was observed when water vapour was present. The ‘water effect’ is the result of a direct reaction of water vapour with the Criegee biradical (the main intermediate in reactions of ozone with alkenes). From our results, we conclude that this reaction could have a more important role in the formation of H_2O_2 and other hydroperoxides than the self- or cross-reactions of peroxy radicals, thereby contributing to forest decline.

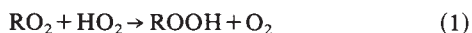
Reaction mixtures were prepared by first injecting the alkene, using a microlitre syringe, into a 130-l glass cylinder reactor. The pressure in the reactor was then brought to atmospheric pressure by adding synthetic air. Ozone, generated by passing pure oxygen through a discharge from a Tesla coil, was then introduced. All gases (Messer Griesheim) and chemicals (Aldrich) were used as supplied. The alkene concentration was monitored before and during the reaction by gas chromatography and the ozone concentration was determined using the ultraviolet absorption at 254 nm. Hydrogen peroxide concentrations were monitored by means of a tunable-diode laser spectrometer (TDLAS) coupled to a 136-m White cell. The TDLAS/White cell system, its performance characteristics and the calibration procedure for trace-gas concentration measurement have been described elsewhere^{11,12}. The gas mixture was sampled into the White cell at regular intervals ($\geq 30\text{ s}$) through a Teflon needle valve and the spectra recorded. Total pressure, temperature and relative humidity in the reactor were monitored by standard techniques. The experimental conditions were as follows: $T=297\text{ K}$, $P_{\text{tot}}=760\text{ torr}$, $[\text{O}_3]_0=2\text{--}60\text{ p.p.m.v.}$,

[Alkene]₀ = 5–35 p.p.m.v., [H₂O] = 0–15 torr (p.p.m.v. is parts per million by volume).

The molar yield of H₂O₂, defined as the ratio of the H₂O₂ concentration to the concentration of the reacted alkene, was found to be in the parts per thousand range for all alkenes investigated. In the presence of water vapour, however, H₂O₂ yields increased significantly for all alkenes. The highest yield (~2%) was measured in the case of limonene at 50% relative humidity (H₂O partial pressure of ~11 torr). The yield of H₂O₂ increased with an increase in the partial pressure of water vapour and was found to be independent of ozone concentration in the range 2–40 p.p.m.v. at an alkene concentration of 30 p.p.m.v., and independent of alkene concentration in the range 6–60 p.p.m.v. at ozone concentration of 40 p.p.m.v. The yields of H₂O₂ for the terpenes and alkenes investigated, with and without water vapour, are summarized in Table 1.

The gas-phase reaction of ozone with terpenes is known to result in a significant production of aerosols in addition to the gaseous products^{13,14}. Lighter alkenes, however, are expected to produce a much smaller amount of aerosols. The enhancement by water vapour of the H₂O₂ yield in the ozonolysis of isobutene and *trans*-2-butene, as well as in that of terpenes, indicates that the water effect is most probably due to a gas-phase reaction and not to heterogeneous processes in the aerosol phase. The contribution of the latter processes cannot, however, be ruled out completely. The rapid appearance of the H₂O₂ signal in the presence of H₂O rules out the possibility that heterogeneous processes at the reactor wall could contribute to the enhancement effect observed.

Our results suggest a direct involvement of water vapour in the mechanism of alkene ozonolysis leading to the production of H₂O₂. The gas-phase reaction of ozone with alkenes is believed, by analogy with the liquid phase¹⁵, to proceed by the addition of ozone to the unsaturated C=C bond forming a molozonide which decomposes rapidly into a carbonyl compound and an initially energy-rich Criegee biradical (R₁R₂COO*)^{9,10}. The 'hot' Criegee biradical is thought to decompose, either directly or after isomerization, into a 'hot' acid, leading to the formation of stable products and free radicals (such as OH, RO, HO₂, RO₂, HCO). The subsequent cross- or self-reactions of the peroxy radicals resulting from the decomposition of the Criegee intermediate can lead to the formation of hydroperoxides



where R is an alkyl group, an acyl group or a hydrogen atom.

Until now, the disproportionation reaction of the hydroperoxy radical (R = H), in particular, has been considered to be the only important reaction producing H₂O₂ in the gas phase. The rate constant of this reaction is enhanced by a factor of two in the presence of 15 torr of water vapour under normal conditions, through the formation of a HO₂·H₂O complex¹⁶. But this enhancement does not account for the large increase in H₂O₂ yield (up to a factor of 10) that we observe in the presence of water vapour.

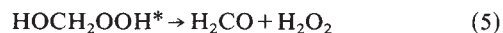
The 'hot' Criegee biradical can also become partly stabilized and react with a number of atmospheric constituents⁹ such as

aldehydes, SO₂, NO_x and H₂O. Of particular interest for the 'water effect' is the reaction of the Criegee biradical with H₂O. This reaction was proposed by Hatakeyama *et al.*¹⁷ to account for the increase in formic acid yield observed when H₂O vapour was added to H₂COO produced by the photolysis of ketene in the presence of O₂. The reaction of water vapour with the Criegee biradical may also explain¹⁸ the formation of peroxyformic acid (HC(O)OOH) observed in the reaction of ozone with chloroethylenes. The possibility of formation of hydrogen peroxide and other hydroperoxides in the gas phase by a similar mechanism has not yet been thoroughly discussed.

This direct reaction between H₂O and the Criegee biradical may account for the observations reported here. The first step of this reaction may be the formation of a R₁R₂COO·H₂O complex, which decomposes into a carbonyl compound and hydrogen peroxide



In the case of the simplest Criegee biradical (H₂COO), reaction (3) is exothermic¹⁹ with an enthalpy of $\Delta H = -128 \text{ kJ mol}^{-1}$. The sequence of reactions (2) and (3) could also proceed through intermediate steps in which the complex rearranges into a 'hot' hydroperoxide which then decomposes (or becomes partly stabilized) into an aldehyde (or a ketone) and hydrogen peroxide, by analogy with the aqueous phase²⁰



A mechanism involving reaction (4) has been previously proposed^{10,21} to account for the observation of hydroxymethylhydroperoxide (HOCH₂OOH) and bis(hydroxymethyl)peroxide (HOCH₂OOCH₂OH) in the gas-phase ozonolysis of some alkenes and terpenes.

To test the possibility of such a direct reaction between water vapour and the Criegee biradical, we conducted two further experiments. In the first, the reactions of ozone with terpenes were reinvestigated under atmospheric conditions replacing H₂O with D₂O. No increase in the H₂O₂ yield was observed in this case. The reaction of D₂O with the Criegee biradical would lead to the formation of D₂O₂ through reactions (2) and (3) and/or HOOD through reactions (4) and (5). Deuterium peroxide (D₂O₂) was identified as reaction product by its absorption spectrum near 948 cm⁻¹. No accurate quantitative determination of D₂O₂ could be carried out at this stage owing to the lack of information on its infrared absorption cross-sections. Preliminary measurements of the infrared cross-sections ($\sigma \leq 10^{-18} \text{ cm}^2$ at a total air pressure of 10 torr), however, indicated that the D₂O₂ concentrations exceeded those of H₂O₂ measured under the same conditions in the absence of D₂O. The formation of D₂O₂ by isotope exchange between H₂O₂ and added D₂O was therefore discounted. Such an isotope exchange would also result in the suppression of H₂O₂, which was not observed. In the case of HOOD, infrared absorption spectra and cross-sections are not available and so positive identification was not possible. In the second experiment, the reaction of ozone and limonene was investigated under 'oxygen-free' conditions in 1 bar of nitrogen. Oxygen-free ozone was obtained by passing the ozone-oxygen mixture from the silent discharge over silica gel in a cold trap at -196 °C and pumping off the excess oxygen. The H₂O₂ concentration was below the detection limit (0.6 parts per 10⁹ by volume) corresponding to a yield upper limit of $\sim 5 \times 10^{-5}$ with 10 p.p.m.v. ozone and 30 p.p.m.v. limonene. In the presence of water vapour (11 torr), however, the H₂O₂ yield increased to ~1% in the oxygen-free system. Under the conditions of the experiments, the formation of H₂O₂ by HO₂ disproportionation should be significantly reduced. The results of these two experiments can be accounted for only by assuming a direct reaction between water vapour and the Criegee biradical.

TABLE 1 Molar yield of H₂O₂ in the reaction of ozone with some alkenes and terpenes

Hydrocarbon	% H ₂ O ₂ (without H ₂ O)	% H ₂ O ₂ (10 torr H ₂ O)
Isoprene	0.036	0.1
β-Pinene	0.04	0.15
α-Pinene	0.10	0.5
Δ ³ -Carene	0.13	0.6
d-Limonene	0.30	1.8
trans-Butene	0.06	0.5
Isobutene	0.04	0.37

Whether the mechanism of H_2O_2 production proceeds by direct formation (reactions (2) and (3)), or by an intermediate hydroxylhydroperoxide (reactions (4) and (5)) requires further investigation.

The reaction of the Criegee biradical with water vapour can be another source for the production of hydrogen peroxide and possibly other peroxides in the gas phase that could be less sensitive to NO_x atmospheric mixing ratios than the peroxy radical reaction (1). The formation of hydroperoxides through reaction (1) depends to a great extent on the ambient NO_x concentrations, NO in particular, which compete with HO_2 for the RO_2 radical. The rates of the reactions of HO_2 with NO ($k = 8 \times 10^{-12} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$) and with RO_2 ($k \leq 6 \times 10^{-12}$) favour the formation of hydroperoxides under conditions of low NO concentrations (< 10 parts per 10^{12} by volume) such as those found in remote unpolluted areas. The concentrations of NO can be orders of magnitude higher in rural and moderately polluted regions. The rate constant of the reaction of NO with the Criegee biradical (H_2COO) has not yet been measured, but is expected to be large ($\sim 10^{-11}$) by analogy with the reaction of NO with HO_2 and CH_3O_2 . For the reaction of the biradical with H_2O , there is a large range of reported values (10^{-19} – 10^{-15}) (ref. 22), the most recently determined value²³ being $1.6 \times 10^{-17} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$. Under conditions of high relative humidity, the reaction with water vapour is therefore expected to be the main loss pathway for the Criegee biradical. Accurate determination of the rate constants of the reactions of the biradical with H_2O and other pollutants, in particular NO , are needed to assess the role of this reaction in the formation of H_2O_2 and other peroxides in the atmosphere. The ozonolysis of naturally occurring alkenes could provide an explanation for the high concentrations of H_2O_2 and other hydroperoxides found in forest areas, especially under conditions of high relative humidity and in plant leaves through the stomatal openings and the inner air space²⁴. Hydrogen peroxide, in particular, has been shown to significantly inhibit plant photosynthesis¹ and to damage the basic constituents of biological systems, such as nucleic and amino acids³. Further experiments are being carried out to elucidate the reaction mechanism responsible for the increase in the yield of hydrogen peroxide when water vapour is present and to obtain quantitative information about the rate constants of the reactions involved in order to assess their importance in forest air chemistry. $\square$

Recent increase in nitrate concentration of Antarctic snow

Paul A. Mayewski* & Michel R. Legrand†

* Glacier Research Group, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA

† Laboratoire de Glaciologie et Géophysique de l'Environnement, BP 96, 38042 St Martin D'Heres Cedex, France

POLAR ice cores provide a unique record of global climate change. In particular, their records of nitrate concentration can yield new insight into the atmospheric nitrogen cycle, but first it is necessary to understand the processes controlling the spatial distribution of nitrate at the ice-sheet surface, and to define any trends in its temporal distribution. Here we report trends in the nitrate time series deduced from low-accumulation sites such as Dome C and Vostok in Antarctica. These trends must be treated with caution because of the possibility of post-depositional alteration. But the increases in the concentration of the spring maximum in nitrate that we find in the South Pole record for the past few years deserve careful consideration, as they may be a result of denitrification of polar stratospheric clouds in the lower stratosphere and may hence be connected in some way with the Antarctic ozone 'hole'.

Nitrate concentrations in snow deposited at several sites in the interior of the Antarctic (Fig. 1) exhibit a marked increase in the upper metre (Fig. 2). The increase appears first in the mid-1970s at Vostok and Dome C and by 1980 in the Dominion Range. The increase appears to be unique and significantly exceeds background values recorded in ice cores at Vostok (based on a continuous profile over the past 34 years (Fig. 2) and a 10-kyr discontinuous time series¹), Dome C (based on a continuous record covering the past 100 years (Fig. 2) and a 10-kyr discontinuous time series²) and the Dominion Range (based on a continuous time series of 84 years (Fig. 2) and a 6-kyr discontinuous time series³).

In December 1988, 6-m snowpit (South Pole T, Fig. 2) was sampled 38 km from South Pole station along the eastern margin of the clean-air sector (to avoid local contamination from the station). The South Pole T snow layers were dated using various independent methods. The well marked volcanic fallout of Mt Agung (1963), as well as radionuclide events (1955 and 1963), were identified and indicated a mean annual accumulation rate of 6.4 – $8.2 \text{ g cm}^{-2} \text{ yr}^{-1}$ over the period of the record (1955–1988)⁴. Detailed year-by-year and seasonal dating was based on the measurement of sodium at a sample interval of $\sim 1.1 \text{ cm}$.

Winter NO_3 values at South Pole T remain fairly constant throughout the record. But the NO_3 maxima which correspond to spring precipitation (as clearly identified by minima in our sodium profile) have been increasing over very recent years (Fig. 2), and there was a particularly large spike in the spring of 1987 (380 ng g^{-1}). These high spring NO_3 values over the past four years significantly exceed background values recorded in South Pole snow layers (except for one event ~ 180 years ago, which is believed to be spurious⁵).

Compared with the mean value of NO_3 at South Pole ($102 \pm 10 \text{ ng g}^{-1}$), Dome C, Vostok and, to a lesser extent, the Dominion Range are characterized by low NO_3 values (before 1975, 19 ± 8 , 16 ± 4 and $59 \pm 12 \text{ ng g}^{-1}$, respectively).

Although a few workers have discussed the HNO_3 budget of Antarctic snow^{5,6} no convincing explanation has been proposed to explain the geographical variation of nitrate over the Antarctic. Such a large difference (by a factor up to five) between Vostok and the South Pole, both located in the central high plateau at similar elevations, is difficult to understand because nitrate is believed to be a product of long-range transport and local sources are not expected to be significant. An explanation for the NO_3 distribution could be that NO_3 is re-emitted to the atmosphere after deposition, reducing the concentration in older

Received 2 April; accepted 29 May 1990.

- Masuch, D., Kettrup, P., Mallant, A. & Slanina, J. *Verein Deutscher Ingenieure-Berichte* **560**, 761–776 (1985).
- Möller, D. *Atmos. Envir.* **23**, 1625–1627 (1989).
- Sies, H. *Angew. Chem.* **98**, 1061–1075 (1986).
- Rennenberg, H. in *Symp. Verteilung und Wirkung von Photooxidanten im Alpenraum*, 360–370. (Gesellschaft für Strahlen und Umweltforschung, Garmisch-Partenkirchen, 1988).
- Hellpöinter, E. & Gáb, S. *Nature* **337**, 631–634 (1989).
- Kok, G. L. & McLaren, S. E. *Proc. Int. Conf. Generation of Oxidants on Regional and Global Scale* abstr. 6–3 (University of East Anglia, 1989).
- Hewitt, N. C., Kok, G. L. & Fall, R. *Nature* **344**, 56–58 (1990).
- Jacob, P. & Klockow, D. *Proc. 3rd French-German Workshop on Tropospheric Chemistry by Laboratory Studies* (eds Becker, K. H. & Carlier, P.) 77–79 (Universitt Wuppertal, 1987).
- Atkinson, R. & Lloyd, A. C. *J. phys. Chem. Ref. Data* **13**, 369–444 (1984).
- Martinez, R. I., Herron, J. T. & Huie, R. E. *J. Am. chem. Soc.* **103**, 3807–3820 (1981).
- Bechara, J., Becker, K. H. & Brockmann, K. J. *Proc. 5th European Symp. Physico-Chemical Behaviour of Atmospheric Pollutants* (Restelli, G. & Angeletti, G.) 27–31 (Kluwer, Dordrecht, 1990).
- Becker, K. H., Brockmann, K. J. & Bechara, J. *Geophys. Res. Lett.* **16**, 1367–1370 (1990).
- Graedel, T. E. *Rev. Geophys. & Space Phys.* **17**, 937–947 (1979).
- Yokouchi, Y. & Ambe, Y. *Atmos. Envir.* **19**, 1271–1276 (1985).
- Bailey, P. S. *Chem. Rev.* **58**, 925–1010 (1958).
- Hamilton, E. J. & Liu, R. *Int. J. chem. Kinet.* **9**, 875–885 (1977).
- Hatakeyama, S., Bandow, H., Okuda, M. & Akimoto, H. *J. phys. Chem.* **85**, 2249–2254 (1981).
- Niki, H., Maker, P. D., Savage, C. M. & Breitenbach, L. P. *J. phys. Chem.* **86**, 1858–1861 (1982).
- Wadt, W. R. & Goddard, W. A. *J. Am. chem. Soc.* **97**, 3004–3021 (1975).
- Yamamoto, Y., Niki, E., Shikawa, H. & Kamiya, Y. *J. org. Chem.* **44**, 2137–2142 (1979).
- Gáb, S., Hellpöinter, E., Turner, W. V. & Korte, F. *Nature* **316**, 535–536 (1985).
- Herron, J. T., Martinez, R. I. & Huie, R. E. *Int. J. chem. Kinet.* **14**, 201–224 (1982).
- Suto, M., Manzanarez, E. R. & Lee, L. C. *Envir. Sci. Technol.* **19**, 815–820 (1985).
- Chameides, W. L. *Envir. Sci. Technol.* **23**, 595–600 (1989).

ACKNOWLEDGEMENTS. K.J.B. thanks W. Seiler for his interest in this work. This research was supported by the Ministerium für Umwelt, Raumordnung und Landwirtschaft (Nordrhein-Westfalen).

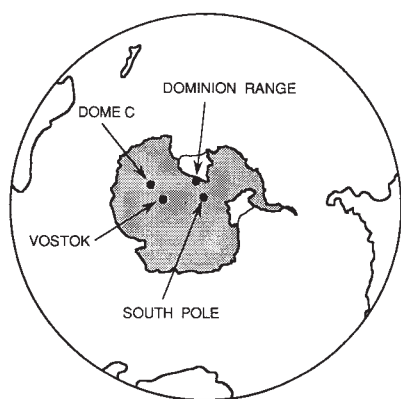


FIG. 1 Map of the Antarctic showing site locations.

snow⁷. This phenomena should be more effective at low-accumulation sites, so Vostok, Dome C and to a lesser extent the Dominion Range (accumulation rates ranging from 2.1 to 3.5 g cm⁻² yr⁻¹) could all be affected by this process. Therefore caution is required in interpreting the NO₃ trend observed at these locations.

At the South Pole such a post-depositional effect can be discounted by comparing our South Pole profile with previously collected records (Fig. 2). The NO₃ concentrations recorded in the upper parts of a core collected in 1981–82 do not display an increase in NO₃ whereas a record collected in 1985 (ref. 8) displayed early signs of the NO₃ increase.

A change in the accumulation rate could modulate the NO₃ concentration in snow. For example, a reduced accumulation rate could enhance NO₃ concentrations if a large amount of NO₃ were deposited directly on the snow surface. We do not believe that such a factor has affected our South Pole record. First, unlike Na and SO₄ dry deposition does not play an important part in the total content of NO₃ in Antarctic snow^{5,7}. Second, we observe no significant change of Na in our profile after the mid-1980s. We therefore assume that the changes in our South Pole T record reflect atmospheric change.

Recent work indicates that major contributions to Antarctic NO₃ budgets could come from lightning, N₂O oxidation and, to a lesser extent, from galactic cosmic rays and surface sources⁵. To explain the dramatic NO₃ rise since the mid-1980s we focus only on contributions that could have changed in the last few decades.

Ice-core studies have established that NO₃ levels in the Northern Hemisphere have doubled since about 1955 owing to contributions from North American and Eurasian anthropogenic emissions^{9,10}. Significant impact of the Northern Hemisphere emissions at high southern latitudes is unlikely because of the short lifetime of NO_x in the lower troposphere¹¹. Biomass burning, together with rising motion in intertropical zones, is probably more effective at supplying NO_x to the Southern Hemisphere free (remote) troposphere.

Biomass burning has increased over the past few centuries and, as of the early 1980s, 70% of it is believed to be a product of human activity¹². Comparing our 1980s NO₃ levels with levels of NO₃ in older ice (covering several decades before the 1980s) it is clear that if any trend exists, it is very weak (10–20 ng g⁻¹). Therefore, the larger NO₃ increase since 1985 could only be due

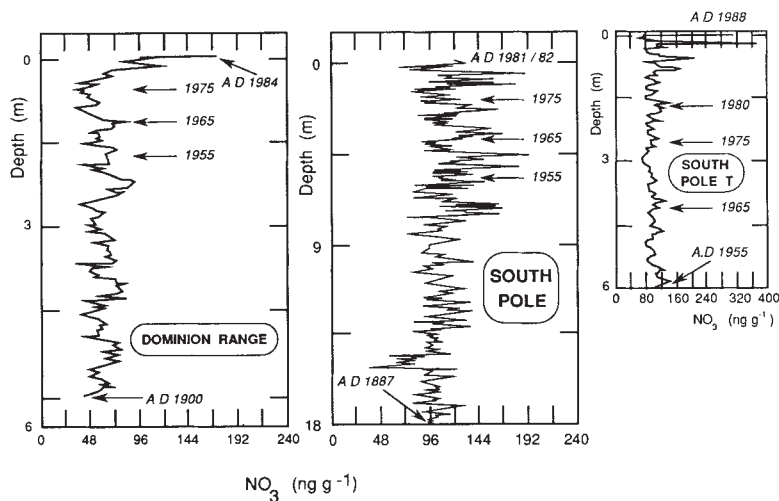
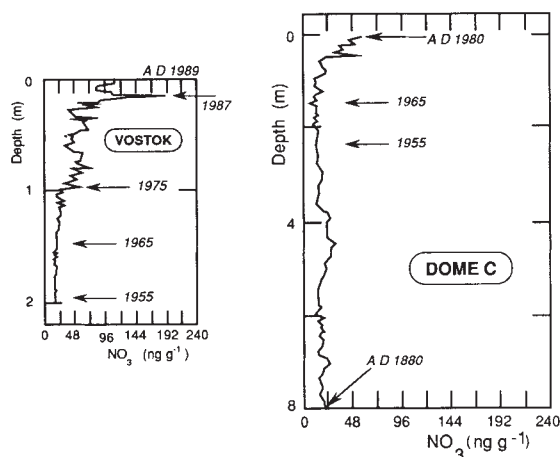


FIG. 2 Time series of the past ~100 years of NO₃ from: Dominion Range, South Pole, South Pole T (see text for location), Vostok and Dome C. Note that the South Pole record shows no increase in NO₃ because the core was collected in 1981–82, whereas the more recently collected South Pole T sequence does record the increase. Similarly, Dome C records only the early stages of the NO₃ increase because the record stops at 1980. The dating of the 1987 spike at Vostok is ± 6 months.



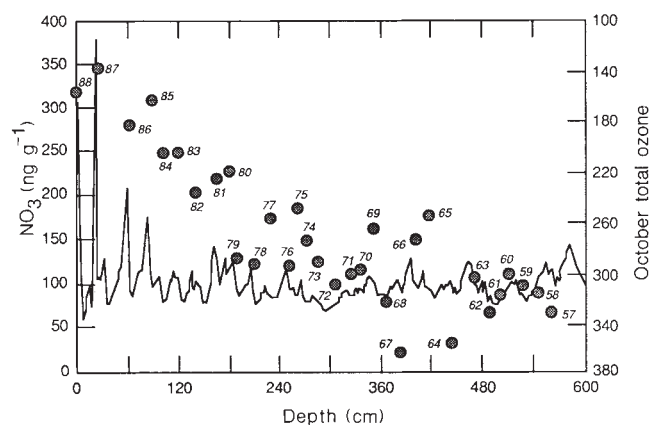


FIG. 3 Detailed NO_3 from South Pole T and October total ozone concentrations (Dobson units), specified by year, based on records from Halley Bay for 1957–63, 1972 and 1983 (October monthly mean)¹⁸ and from the South Pole for all other years (15–31 October mean)^{19,20}.

to biomass burning by humans if these emissions have drastically increased over very recent years because there is little evidence in the NO_3 record for earlier effects of biomass burning. Although some examples of such a dramatic change exist in several locations (such as Rondonia in Brazil where the amount of deforestation, associated with biomass burning, increased by a factor of four to five between 1980 and 1985 (ref. 13)), it is unlikely that this change is representative globally.

Polar stratospheric clouds (PSCs), containing sequestered HNO_3 , form in late June and July. Over the course of the winter (July–October) their altitude decreases (from 25 to 15 km)¹⁴. The subsequent increase of tropospheric HNO_3 level could be reflected in spring polar precipitation (during or after October) and the intensity of this process may be modulated by the amount and persistence of PSCs and, therefore, by the temperature of the polar vortex. Although past trends in PSCs remain unknown, several changes are expected to have possibly enhanced PSC occurrence over the past few decades (such as an increase in methane, which provides more water as an oxidation by-product¹⁵, long-term sea surface temperature changing the tropospheric forcing of the stratosphere¹⁶). Although no clear trend over the past decade has been identified in September PSCs¹⁷, a marked increase in October PSC sightings over the 1979–89 decade in the Quasi-Biennial Oscillation in westerly phase years is reported¹⁷, which mainly reflect decreasing heating rates (through a decrease in ozone). A particularly good example of such a phenomena is the more frequent and persistent PSCs reported in the spring of 1987.

Denitrification of the lower stratosphere could be the link between the recent NO_3 increases that we observe and the austral polar ozone loss occurring during September and October. A comparison of the October total ozone values with the NO_3 measured at South Pole T (Fig. 3) reveals that the four largest ozone depletions (1985, 1986, 1987 and 1988) are all associated with the high NO_3 maxima and the NO_3 maxima for the three most recent years are the highest for at least the past 100 years.

Our South Pole NO_3 record has a very large peak (380 ng g^{-1}) in the spring of 1987. Although the Vostok data must be interpreted with caution, as discussed above, this profile also has a large spike (190 ng g^{-1}) near the surface (Fig. 2) dated at 1987 ± 6 months.

Our data suggest that since 1987, and possibly since 1985, more persistent PSCs and subsequently greater removal of NO_3 from the lower stratosphere to the troposphere disturbed the NO_3 budget of the troposphere producing the increase in NO_3 deposited at the South Pole, and perhaps elsewhere. Although it is still difficult to claim that the recent change in NO_3 deposition is a consequence of the spring ozone depletion, because

the response is not absolutely direct, we suggest that the link may be through the relationship between PSCs and O_3 depletion.

Further detailed examination of the chemical composition of snow deposited at the South Pole during the spring of 1989, the largest ozone depletion to date, could provide a test of the putative link between ice-core NO_3 , PSCs and O_3 . □

Received 2 April; accepted 22 May 1990.

1. Legrand, M. R., Lorius, C., Barkov, N. I. & Petrov, V. M. *Atmos. Environ.* **22**, 317–331 (1988).
2. Legrand, M. R. thesis, CNRS Grenoble (1985).
3. Mayewski, P. A., Spencer, M. J., Lyons, W. B., Twickler, M. S. & Dibb, J. E. *Antarct. J. U.S.* **23**, 64–68 (1988).
4. Dibb, J. E., Mayewski, P. A., Buck, C. F. & Drummey, S. M. *Nature* **345**, 25 (1990).
5. Legrand, M. R. & Kirchner, S. *J. geophys. Res.* **95**, 3493–3509 (1990).
6. Legrand, M. R. & Delmas, R. *Tellus* **B38**, 236–249 (1986).
7. Neubauer, J. & Heumann, K. G. *Atmos. Environ.* **22**, 537–545 (1988).
8. Laird, C. M. thesis, Univ. Kansas (1986).
9. Neftel, A., Beer, J., J., Oeschger, H., Zurcher, F. & Finkel, R. C. *Nature* **314**, 611–613 (1985).
10. Mayewski, P. A. *et al. Science* **243**, 975–977 (1986).
11. Ehhalt, D. & Drummond, J. W. in *The Tropospheric Cycle of NO_3 , Chemistry of the Unpolluted and Polluted Troposphere* (eds Georgii, W. H. & Jaeschke, W.) 219–251 (Reidel, Dordrecht, 1982).
12. Ehhalt, D. & Drummond, J. W. in *Tropospheric Ozone* (ed. Isaksen, I. S. A.) 217–237 (Reidel, Dordrecht, 1988).
13. Malingreau, J. P. & Compton, J. T. *Ambio* **17**, 49–55 (1988).
14. McCormick, M. P. & Trepte, C. R. *Geophys. Res. Lett.* **13**, 1276–1279 (1986).
15. Blake, D. R. & Rowland, F. S. *Science* **239**, 1129–1131 (1988).
16. Schoeberl, M. R. & Krueger, A. J. *Geophys. Res. Lett.* **13**, 1191–1192 (1986).
17. Poole, L. R., Solomon, S., McCormick, M. P. & Pitts, M. C. *Geophys. Res. Lett.* **16**, 1157–1160 (1989).
18. Watson, R. T. *et al. in NASA Ref. Publ.* 1208, 88 (1988).
19. Kornyr, W. D., Grass, R. D. & Reitelbach, P. L. *J. geophys. Res.* **84**, 11429–11436 (1989).
20. Krueger, A. J., Stolarski, R. S. & Schoeberl, M. R. *Geophys. Res. Lett.* **16**, 381–384 (1989).

ACKNOWLEDGEMENTS. We thank B. Lopez, M. Morrison, J. R. Petit, C. Wake and the staff at South Pole for their help in the field, C. F. Buck, M. J. Spencer and S. Whitlow for their work in the laboratory, J. Dibb for providing data used to assign chronologies, and J. Dibb and D. Skole for comments. This research was supported by the NSF and le Centre National de la Recherche Scientifique.

Limited diffusion of U-series radionuclides at depth in deep-sea sediments

S. Colley & J. Thomson

Institute of Oceanographic Sciences, Deacon Laboratory, Brook Road, Wormley, Godalming GU8 5UB, UK

LONG-TERM radionuclide mobility is a primary concern for evaluating options for the disposal of radioactive waste. Information relevant to the long timescales necessary may be obtained from natural radionuclides. Here we examine a characteristic irregular profile of uranium concentration with depth, in deep-sea turbidite units more than 500,000 years old. This is enough time either for secular equilibrium to have been established, or for daughter radionuclides to have achieved steady-state distributions in response to local advection and diffusion. We find that advection is absent and that the ^{238}U -series radionuclides ^{234}U and ^{230}Th are remarkably immobile, with very small effective diffusion coefficients, $D < 5 \times 10^{-13} \text{ cm}^2 \text{ s}^{-1}$. But a fraction of the total ^{226}Ra (another daughter radionuclide in the ^{238}U decay series) is mobile, with $D = 10^{-9} - 10^{-10} \text{ cm}^2 \text{ s}^{-1}$. This work suggests that the reducing nature of the sediments at depth is important in maintaining such low D values. The *in situ* diffusion behaviour observed must be correctly predicted by any model used for evaluating waste disposal options.

Current suggestions for deep-sea disposal of high-level radioactive waste envisage packaging the vitrified waste in steel canisters and burying the canisters in sediments^{1,2}. Waste radionuclides would only be released to the ocean after canister corrosion, glass leaching, and finally migration through the few tens of metres of overlying sediments (the 'sediment barrier') had occurred. Some important waste nuclides have long half-lives so isolation times from the biosphere of hundreds of thousands of years are desired. Forecasts of behaviour on such timescales are made by mathematical models, and involve large

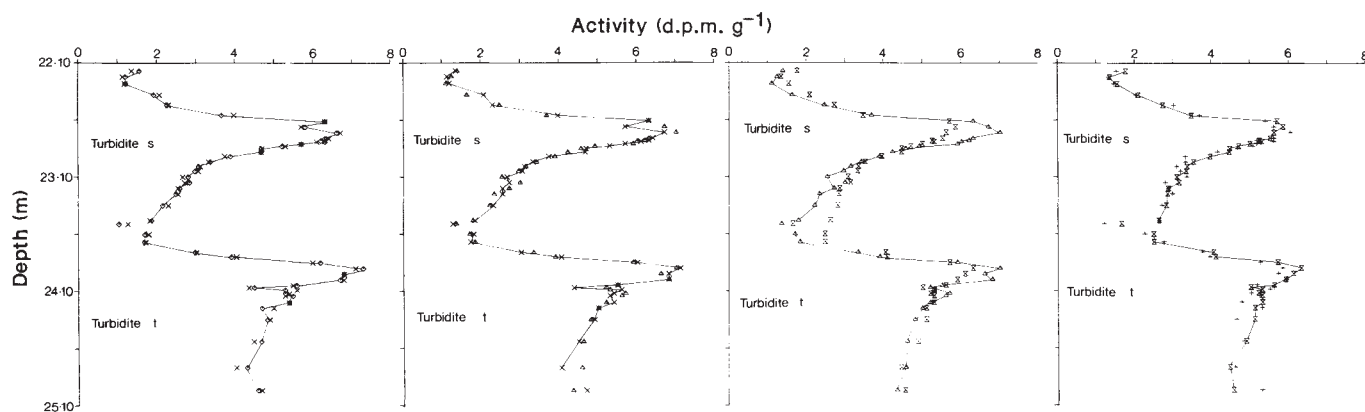


FIG. 1 Activity depth profiles for parent-daughter pairs in the ^{238}U series. The daughter radionuclide in each panel becomes the parent radionuclide in the next panel and retains the same symbol. From left to right ^{238}U (diamonds)– ^{234}U (diagonal crosses); ^{234}U – ^{230}Th (triangles); ^{230}Th – ^{226}Ra

(bow-ties); ^{226}Ra – ^{210}Po (rectangular crosses). In each panel, the parent profile points are joined within each of the two turbidites. 1σ counting errors (not shown) are generally 1.5–4% for ^{238}U , ^{234}U and ^{230}Th , 1–2% for ^{226}Ra and 3–5% for ^{210}Po .

extrapolations from any input parameters derived from experiments on laboratory timescales. Thus the study of natural geochemical localizations of radioelements that have formed and persisted in the environment are valuable to obtain data relevant for long times. Studies of the natural decay series based on primordial ^{238}U , ^{235}U and ^{232}Th are particularly pertinent because artificial actinides decay through the same chains of radioelements as these natural actinides.

A characteristic uranium depth profile from organic-rich deep-sea turbidites has been identified and investigated^{3–5}. The profile is the result of the turbidites having a relatively constant uranium content throughout on emplacement, followed by redistribution of authigenic uranium near the tops of the units as a result of bottom-water oxidation. This oxidized uranium fraction diffuses from the top of the unit and becomes immobilized by reduction in the immediately underlying anoxic sediments. The redistribution ceases on emplacement of the next turbidite, and thereafter the uranium profile shape has been observed to persist for at least 750 kyr⁵. Previous work has concentrated on the geochemical behaviour of uranium, but here we focus on the daughter radionuclides produced from ^{238}U .

The established stratigraphy of the central Madeira Abyssal Plain⁶ allows a good estimate of the ages of different turbidite units. The times of preservation of the uranium profiles are therefore known and can be used to calculate the growth of the ^{238}U daughter radionuclides. Given sufficient time and no loss or migration of any member of the decay chain, secular equilibrium, where the activity of each chain member is equal to that of ^{238}U , will be established at all depths. The daughter activity profiles relative to that of ^{238}U in a sufficiently old sediment

section therefore represent equilibria in the column with respect to radioactive production, diffusion and advection. It is such sections (>500 kyr old) that we have investigated.

Sections of core MD24, a 30-m piston core collected on the ESOPE (Étude des Sédiments Océaniques par Pénétration) cruise⁷ in 1985, were selected and analysed by standard radiochemical methods for the longer-lived daughters in the ^{238}U decay series (^{234}U , ^{230}Th , ^{226}Ra and ^{210}Po as a proxy for ^{210}Pb)⁸. These are the important chain members; other daughters persist for too short a time to diffuse appreciably. Sampling was

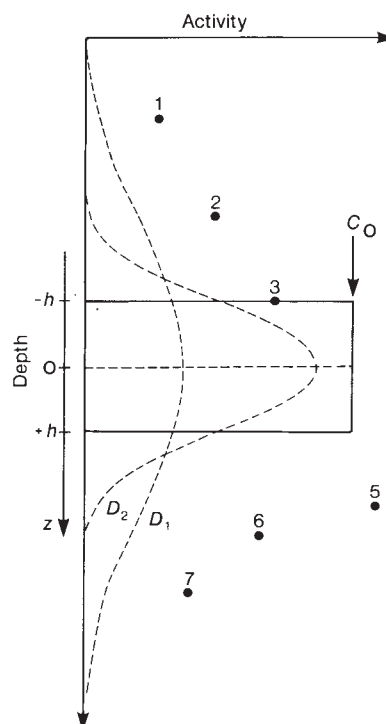


FIG. 2 A sketch showing how a single element contributes to a daughter profile. Points 1–7 are experimental parent data points determined at different depths. The rectangular strip shown is the average of the parent data between points 3 and 4. The two curves demonstrate steady-state daughter distributions developed over the daughter mean life for two different values of D ($D_1 > D_2$). The contributions from the strip are evaluated at depths 1–7. In the model, a single D value is selected and the corresponding daughter profile constructed by summation of the contributions from all strips at all depths. These model daughter values are compared with the experimental daughter values from the same depths. The best fit is achieved by varying D to minimize the SSDs.

TABLE 1 Isotope data and activity ratios

	Turbidite s*	Turbidite t†
$\Sigma ^{238}\text{U}$ (d.p.m. g^{-1})	97.7 ± 0.5	92.1 ± 0.7
$\Sigma ^{234}\text{U}$ (d.p.m. g^{-1})	97.8 ± 0.5	92.0 ± 0.7
$\Sigma ^{230}\text{Th}$ (d.p.m. g^{-1})	97.7 ± 0.5	92.9 ± 0.6
$\Sigma ^{226}\text{Ra}$ (d.p.m. g^{-1})	96.3 ± 0.3	92.9 ± 0.4
$\Sigma ^{210}\text{Po}$ (d.p.m. g^{-1})	95.7 ± 0.6	91.2 ± 0.6
$^{234}\text{U}/^{238}\text{U}$	0.997 ± 0.004	0.999 ± 0.005
$^{230}\text{Th}/^{234}\text{U}$	0.976 ± 0.006	1.004 ± 0.008
$^{226}\text{Ra}/^{230}\text{Th}$	0.975 ± 0.006	0.986 ± 0.007
$^{210}\text{Po}/^{226}\text{Ra}$	0.988 ± 0.007	0.980 ± 0.007

The 1σ uncertainties quoted were derived from the counting statistics.

* 26 analyses.

† 19 analyses.

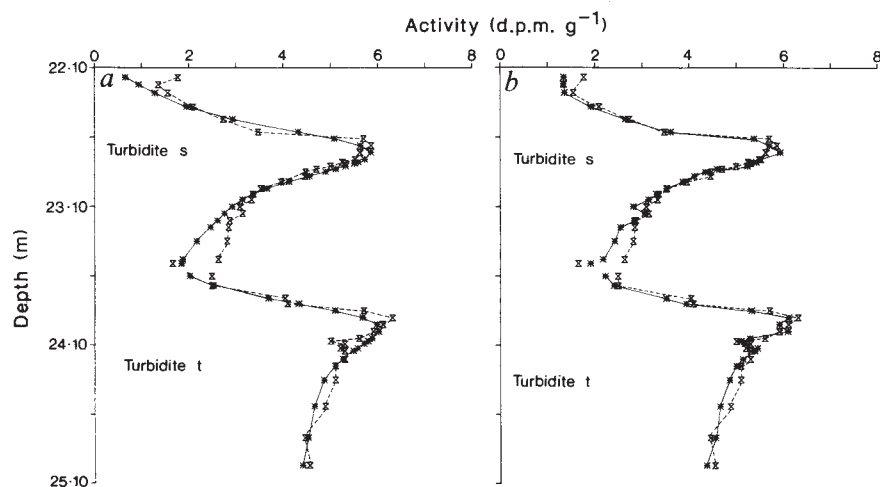


FIG. 3 Best fits between model ^{226}Ra (stars) and experimental ^{226}Ra (bow-ties) profiles for turbidites s and t. Left, 100% ^{226}Ra assumed mobile, with $D = 7 \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$. Right, 35% ^{226}Ra assumed mobile with $D = 6 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$.

concentrated on organic-rich turbidite units because they provide long sections with relatively uniform physical and chemical compositions, but irregular uranium profiles. The clearest evidence of radioactive daughter mobility is expected where the uranium content changes rapidly with depth.

Figure 1 illustrates the data obtained for parent-daughter pairs in a core section, comprising two turbidite units, s and t. It is evident that there is little daughter migration except for ^{226}Ra . No preferential upwards or downwards displacement of any daughter is evident, confirming earlier conclusions⁹ that advection is absent from this area of the sea floor. The system seems to be in secular equilibrium, as demonstrated by the isotope summations (Table 1). To quantify diffusion, a model was developed⁸ that treats the data in parent-daughter pairs. The parent profile is considered as stationary and continuously generates a daughter which is free to migrate with the chosen effective diffusion coefficient D for a period of the daughter mean life. The model assimilates the parent data as a series of strips, each of which behaves according to¹⁰

$$C(z, t) = \frac{C_0}{2} \{ \text{erf} [(h - z)/2\sqrt{Dt}] + \text{erf} [(h + z)/2\sqrt{Dt}] \}$$

where C_0 is the initial activity of the strip, $2h$ is the strip width, t is the mean life for the daughter nuclide, and C is the activity of the daughter nuclide at distance z from the strip centre (Fig. 2). The modelled daughter profile, assuming the same D for all strips, is obtained by summation of the contributions of each strip at the original sample depths. The fit between the model and experimental daughter data is then assessed by minimizing the sum of the squares of the differences (SSD) of the two profiles.

No diffusion could be detected with this model for the pairs ^{238}U – ^{234}U and ^{234}U – ^{230}Th . Decreasing D values yield decreasing SSD values until a constant SSD, equal to that of the input experimental parent-daughter data, is reached at $D \leq 10^{-13} \text{ cm}^2 \text{ s}^{-1}$. ^{230}Th is not expected to be mobile in sediments¹¹, but previous work on deep-sea sediments^{12,13} and marine minerals¹⁴ has shown that a fraction of ^{234}U is usually more mobile than ^{238}U . This is true of sediments in both oxic and anoxic environments, and generally ~30% of the total ^{234}U seems to be mobile with a D value of 10^{-8} – $10^{-9} \text{ cm}^2 \text{ s}^{-1}$ (refs 12, 14). Such behaviour is ascribed to release owing to radiation damage of ^{234}U to solution in the U(VI) oxidation state, following production from ^{238}U present as U(IV). Experiments performed on the ESOPE sediments demonstrated that ^{234}U could indeed be released preferentially, with $^{234}\text{U}/^{238}\text{U}$ activity ratios between 1.08 and 1.84 released to solution by mild chemical leaching⁸. The near-perfect concordance of the ^{234}U and ^{238}U activity summations (Table 1) and $^{234}\text{U}/^{238}\text{U}$ activity ratios close to the secular equilibrium value of 1.00 (Fig. 1), however, show

that the implied potential for ^{234}U mobility must be suppressed under the mildly reducing (post-oxic) conditions in these sediments⁷. This is consistent with the findings of Santschi *et al.*¹⁵ who found much lower uranium concentrations in pore water in core MD24 than those expected on thermodynamic grounds.

^{226}Ra is clearly deficient where ^{230}Th is at maximum values, and in excess at ^{230}Th minima (Fig. 1). If all the ^{226}Ra present is free to diffuse, the model best reproduces the data for the core section of Fig. 1 with $D = 7 \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$ (Fig. 3). Other work concerned with recent near-surface sediments, however, has demonstrated that only a fraction of the total ^{226}Ra can diffuse^{16–18}. Values of 0.35–0.75 for this mobile fraction have been estimated by a ^{222}Rn emanation method^{16,17}, with the lower values observed in carbonate-rich sediments like those used here. We have used both modelling and experimental methods to estimate the mobile fraction. By adapting the model to allow only a fraction of the ^{226}Ra produced from ^{230}Th to diffuse, the best fit to the observed data for turbidites s and t was obtained with a mobile fraction of 0.35 and $D = 6 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$ (Fig. 3). In the experiment, powdered samples with a known range of ^{226}Ra contents were stored in water in sealed bubbler vessels, and the amount of ^{222}Rn available was removed at known times and measured. This experiment also demonstrated that 35% of the ^{222}Rn generated by the total ^{226}Ra present was released. Values of D for ^{226}Ra in the range 1 – $5 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$ with 0.5–0.7 of total ^{226}Ra mobile have been observed by Kadko *et al.*¹⁸ in cases where sorption by manganese oxyhydroxides is not the controlling factor.

The most likely chain member to exhibit mobility in the decay sequence between ^{226}Ra and ^{210}Po is ^{222}Rn . This inert gas is expected to have no interaction with the solid phase, but its short mean life of 5.5 days limits the distance over which it can diffuse. No diffusion could be detected in the fit of the model to the experimental ^{210}Po data.

Turbidites s and t (Fig. 1) are organic-rich and contain both detrital and authigenic uranium components. Only the authigenic uranium is susceptible to the early oxidation process which causes the characteristic profiles. This authigenic fraction is also susceptible to mild leaching (that used earlier to demonstrate differential ^{234}U and ^{238}U reactivity), even in the 500–600 kyr old sediments studied here. Thus, although this work is concerned with natural levels of uranium, not all of the uranium is immobilized by incorporation in mineral lattices. Rather, it is the geochemical conditions of the sedimentary environment that immobilize the uranium and its radioactive daughters. A similar effect would be expected outside the immediate environs of radioactive waste disposed at depth (the ‘far field’), although the magnitude of this immobilizing capacity remains to be determined. □

Received 21 March; accepted 23 May 1990.

- Francis, T. J. G., Searle, R. C. & Wilson, T. R. S. *Phil. Trans. R. Soc. A* **319**, 115–137 (1986).
- Freeman, T. J., Murray, C. N., Francis, T. J. G., McPhail, S. D. & Schultheiss, P. J. *Nature* **310**, 130–133 (1984).
- Colley, S., Thomson, J., Wilson, T. R. S. & Higgs, N. C. *Geochim. cosmochim. Acta* **48**, 1223–1224 (1984).
- Colley, S. & Thomson, J. *Geochim. cosmochim. Acta* **49**, 2339–2348 (1985).
- Colley, S., Thomson, J. & Toole, J. *Geochim. cosmochim. Acta* **53**, 1223–1234 (1989).
- Weaver, P. E. & Rothwell, R. G. in *Geology and Geochemistry of Abyssal Plains*, *Geol. Soc. Spec. Publ.* **31**, 71–86 (1987).
- Schuttenhelm, R. T. E. et al., *Joint Res. Centre Rep.*, EUR 12330 & EUR 12331 (Commission of the European Communities, Luxembourg, 1989).
- Colley, S. & Thomson, J. *Inst. Oceanogr. Sci. Rep.* (in the press).
- Schultheiss, P. J. & Noel, M. in *Geology and Geochemistry of Abyssal Plains*, *Geol. Soc. Spec. Publ.* **31**, 113–129 (1987).
- Crank, J. *The Mathematics of Diffusion* (Clarendon Press, Oxford, 1956).
- Ku, T.-L. *A. Rev. Earth planet. Sci.* **4**, 347–379 (1976).
- Ku, T.-L. *J. geophys. Res.* **70**, 3457–3474 (1965).
- Mangini, A. & Dominik, J. *Sedim. Geol.* **23**, 113–125 (1979).
- Kolodny, Y. & Kaplan, I. R. *Geochim. cosmochim. Acta* **34**, 3–24 (1970).
- Santschi, P. H. et al. *Nature* **331**, 155–157 (1988).
- Cochran, J. K. & Krishnaswami, S. *Am. J. Sci.* **280**, 849–889 (1980).
- Key, R. M., Guinasso, N. L. & Schink, D. R. *Mar. Chem.* **7**, 221–250 (1979).
- Kadko, D., Cochran, J. K. & Lyle, M. *Geochim. cosmochim. Acta* **51**, 1613–1623 (1987).

ACKNOWLEDGEMENTS. We thank M. Talbot for work on the model. This work was carried out under contract to the Department of the Environment and to the Commission of the European Communities as part of their research programmes for radioactive waste disposal.

Evidence for late Proterozoic subduction from 700-Myr-old blueschists in China

T. Nakajima*, S. Maruyama†, S. Uchiumi*, J. G. Liou‡, X. Wang‡, X. Xiao§ & S. A. Graham‡

*Geochemistry Department, Geological Survey of Japan, Tsukuba 305, Japan

†Department of Earth Science and Astronomy, University of Tokyo, Tokyo 153, Japan

‡Department of Geology, Stanford University, Stanford, California 94305, USA

§Chinese Academy of Geological Sciences, Beijing, China

IN a comprehensive review of the world's blueschists, Ernst¹ noted an apparent absence of these high-pressure, low-temperature rocks in the Precambrian, and ascribed it to a cooling of the Earth in the Phanerozoic eon. As blueschists are thought to be formed in subduction zones, the absence of Precambrian blueschists has been adduced as evidence that plate tectonics did not operate on the early Earth². Although this conclusion is not inevitable, in that ancient mobile belts are more likely to have been obliterated than younger ones³, the presence of paired metamorphic belts of kyanite–sillimanite and andalusite–sillimanite grade in Proterozoic Gondwanaland supports the existence of a subducted slab, but with ancient geothermal gradients too high to form blueschists along the convergent plate boundary^{4,5}. On the other hand, data from inclusions in African diamonds support the opposing view, that Archaean continental geotherms were similar to those of today⁶. Here we present Rb–Sr and K–Ar ages of 698–718 Myr for phengitic mica from high-pressure pelitic schists from the Aksu region of China, which are interleaved with blueschists. We interpret these dates as representing the time of the blueschist-facies

metamorphism, suggesting that subduction was occurring in the late Proterozoic.

Possible occurrences in the USSR and China of Precambrian blueschists have been reported^{7–13}. Unfortunately these are not well documented because the data from the isotope analyses and the geological and petrographic descriptions were not provided. Although radiometric ages of 1,600, 1,100 and 600–800 Myr have been obtained for blueschists^{9,10} the orogenic belts in which they occur are of Palaeozoic or even Mesozoic age. Also radiometric ages of blueschists from the same regions cluster around 400–500 Myr. The possible occurrence of Precambrian blueschists is therefore still a subject of debate.

The 15 × 40 km Aksu high-pressure schist terrain, exposed in the northwest margin of the Tarim craton in the Kalpin uplift, is overlain by Recent cover to the north and unconformably overlain by unmetamorphosed upper Proterozoic Sinian sedimentary rocks to the south¹⁴ (Fig. 1). The lowermost Sinian sequence contains some blocks and pebbles of the underlying high-pressure schists. This strongly indicates that the age of the underlying high-pressure schists are older than late Sinian, but until now their isotopic ages had not been determined.

The Aksu schist terrain is composed of stratigraphically coherent pelitic, psammitic, mafic and quartzose schists, and lacks melange facies. It has experienced multi-stage deformation and regional metamorphism of high-pressure intermediate type¹⁵. The metamorphic grade slightly increases towards the north, across the general structure of the terrain. The schist terrain was divided into two mineral zones—winchite and Na-amphibole, based on the occurrences of Na–Ca amphiboles¹⁵; winchite is confined to the winchite zone, whereas coexisting actinolite and Na-amphibole occur in the Na-amphibole zone. Both of the zones correspond to the upper pumpellyite-actinolite facies or high-pressure greenschist-to-blueschist facies. Typical mineral assemblages are: Na-amphibole (or winchite) + epidote + chlorite + albite + quartz ± actinolite in blueschists and phengitic mica + chlorite + albite + quartz in pelitic schists. Considering the similar occurrences of Na-amphiboles and winchites in other metamorphic terrain of high-pressure intermediate type^{16,17} and geothermometry on the equivalent paragenesis¹⁸, the metamorphic temperature of the Na-amphibole zone and the winchite zone are estimated to be ~300 and 350 °C, respectively, with pressure in the range 5–7 kbar. Although the mineral assemblage of pelitic schist by itself is not incontrovertible evidence of high-pressure metamorphic conditions, the intimate interlayering on a centimetre scale of pelitic schists and blueschists with concordant microfolding shows that pelitic schists were also formed by blueschist facies metamorphism.

We determined the ages of two micaceous schists samples, Ak76 and Ak102A. Both samples were collected from the Na-amphibole zone of the Aksu blueschist terrain, ~2 km apart (Fig. 1). The mineral assemblages of these two rocks are the same as those listed above for pelitic schists except for the accessory phases. The difference of metamorphic temperatures between these two rocks is too small to detect, judging from petrological observations and field occurrences. In these rocks, phengitic micas and chlorites form the schistosity, and there are no signs of secondary alteration or mineralization. This indicates

TABLE 1 Isotope data

		Rb (p.p.m.)	Sr (p.p.m.)	⁸⁷ Rb/ ⁸⁶ Sr	⁸⁷ Sr/ ⁸⁶ Sr	Age (Myr)	K ₂ O (%)	Radiogenic ⁴⁰ Ar (×10 ⁻⁶ ml g ⁻¹)	Atmospheric ⁴⁰ Ar (%)	Age (Myr)
Ak76	phengitic mica	347.9	163	6.176	0.77076	698 ± 26	9.59	273	1.99	718 ± 22
	whole rock	103.5	272.4	1.009	0.72020					
Ak102A	phengitic mica	360	65.7	15.85	0.87150	714 ± 24	10.50	295	1.58	710 ± 21
	whole rock	58.8	100.0	1.701	0.72735					

The decay constant of ⁸⁷Rb used for calculating the age was 1.42 × 10⁻¹¹ yr⁻¹ (ref. 24). For K–Ar ages, the decay constants of ⁴⁰K used were 4.962 × 10⁻¹⁰ yr⁻¹ for β-capture and 0.581 × 10⁻¹⁰ yr⁻¹ for electron capture (ref. 24). The uncertainties in the ages are quoted at the 1σ level.

that the phengitic micas in these rocks were formed during metamorphism, and are not detrital. For age determination, we separated phengitic mica from the rocks using a magnetic separator, heavy liquid, and tapping on a sheet of paper.

The Rb-Sr and K-Ar ages were determined using VG-54E and VG-6 mass spectrometers, respectively, at the Geological Survey of Japan. All Rb and Sr concentrations were measured using the isotope dilution method and the ages obtained from whole rock/mineral isochrons. Representative analyses and the calculated ages are shown in Table 1. The phengitic mica from the Aksu high-pressure schists gave K-Ar and Rb-Sr ages of 698–718 Myr. Considering the errors, the age differences between Ak76 and Ak102A are too small to have geological meaning. The closure temperature of the K-Ar system in muscovite is 350 °C, whereas that of the Rb-Sr system is 500 °C, at a cooling rate of 30 °C Myr⁻¹ (refs 19, 20). The petrological features indicate that the metamorphic temperatures of these rocks are nearly at the closure temperature of the K-Ar system in muscovite. The approximate coincidence of K-Ar and Rb-Sr ages of phengitic mica in these rocks implies that these ages approximately reflect the timing of peak metamorphism. Thus, it can be concluded that the high-pressure metamorphism of Aksu blueschists occurred around 698–718 Myr ago, in late Proterozoic time. This result does not contradict the geological observations that have been reported^{14,15}. Moreover, this relationship provides a good time marker in late Precambrian to Cambrian geological history of China.

These isotopic ages imply a late Proterozoic tectonic event in and around the Tarim craton, which involves formation of the high-pressure metamorphic terrain and the deposition of Sinian sedimentary rocks. The absolute age of the lowermost Sinian sequence is still unknown, because of the absence of isotopic age data and a lack of reliable biostratigraphic time markers²¹. In the Yangtze area in Hubei Province, where the most reliable chronostratigraphic sequence has been established, the isotopic ages of the Sinian are mainly in the range 614–740 Myr, whereas

those of the pre-Sinian units are older than 800 Myr²². The absolute age of the oldest Sinian rocks has not yet been established. The coupled K-Ar and Rb-Sr ages reported herein strongly indicate that the age of the lowermost Sinian sequence in the northern Tarim craton is younger than 700 Myr. The late Precambrian age of these blueschists indicates that at least the last phase (700–500 Myr) of the Pan-African orogeny was driven by modern-type plate movements.

In Aksu blueschists, one rarely finds the metamorphic textures formed by pressure-released retrograde reactions, such as zoned amphiboles having a crossite core with an actinolite or riebeckite rim. Based on his compilation of blueschist belts, Ernst²³ has classified blueschists into two categories, those with and those without pressure-released retrograde metamorphic textures. According to him, the former are generated by orogeny of the continental-collision type, typified by the Alps, whereas the latter are formed by Cordilleran-type orogeny, such as that which generated the Franciscan complex in California. If this is correct, the Aksu blueschist metamorphism was caused by subduction of an oceanic plate before the deposition of the Sinian sequence. Inferred protoliths of accretionary-complex affinity, rather than passive-margin sediments, support the subduction-related process. This 700-Myr subduction and associated low geothermal gradient formed blueschists of the high-pressure intermediate type, but not metamorphic rocks of kyanite-sillimanite type. On the basis of our data, it seems unlikely that the high geothermal gradient necessary to form the paired metamorphic belts of kyanite-sillimanite and andalusite-sillimanite types^{4,5} was typical of late Proterozoic subduction zones.

Two possible tectonic settings for the formation of the Aksu blueschist terrain are: (1) The Aksu blueschist terrain was an accretionary complex formed 700 Myr ago along the northern margin of the proto-Tarim craton, which presumably constituted the northernmost front of the Gondwana supercontinent at that time, or (2) it was formed along the southern margin of another

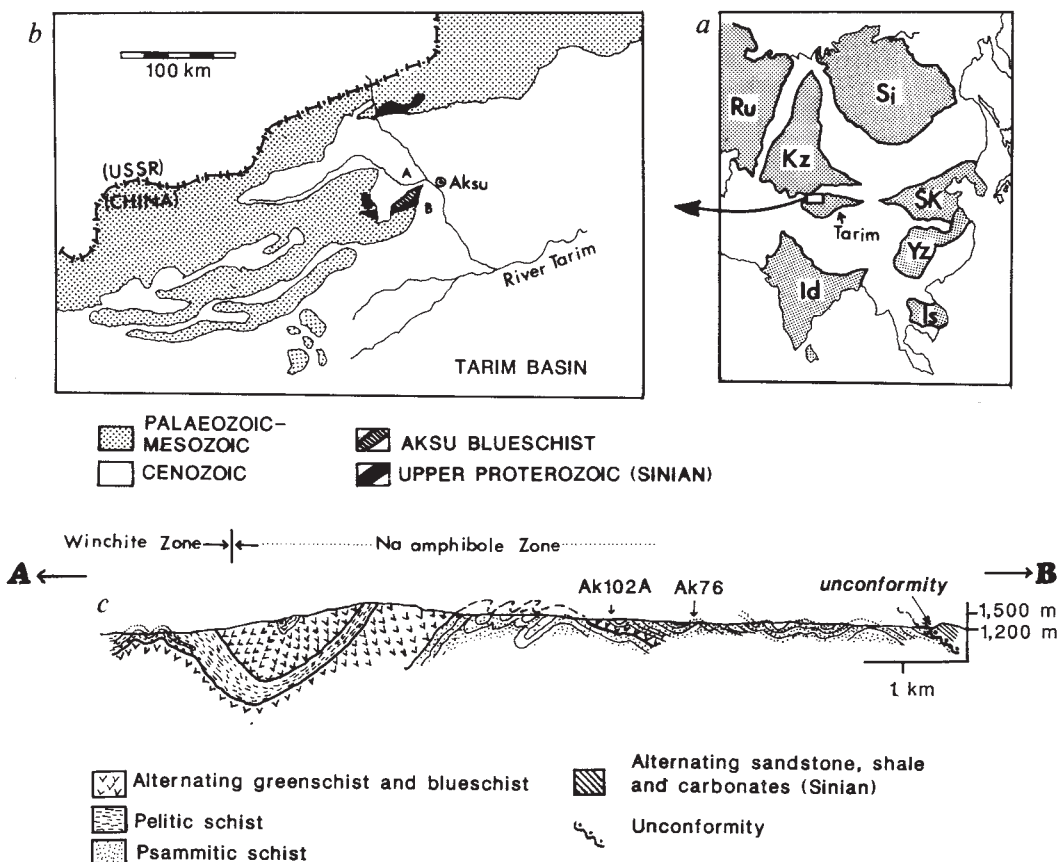


FIG. 1 a, Ancient cratons in Asia. Ru, Russian; Kz, Kazakhstan; Si, Siberian; SK, Sino-Korean; Yz, Yangtze; Id, Indian; Is, Indosinian. b, Geological map of Aksu area. c, Cross-section of the Aksu blueschist terrain along A-B in b showing the sample locations.

unknown craton to the north, and a succeeding rift took place at the southern periphery of that northern craton and its southern offset including the Aksu blueschist terrain adjoined the proto-Tarim craton. The former indicates southward subduction and the latter northward subduction. Whatever the direction of subduction the evolution of the Tarim craton involved a high-pressure low-temperature subduction metamorphism at 700 Myr. □

Received 13 February; accepted 23 May 1990.

- Ernst, W. G. *Am. J. Sci.* **274**, 657–668 (1972).
- Kröner, A. *Precamb. Res.* **4**, 163–213 (1977).
- Draper, G. & Bone, R. *J. Geol.* **89**, 601–613 (1981).
- Katz, M. B. *Geology* **2**, 237–241 (1974).
- Shiraishi, K., Hiroi, Y., Motoyoshi, Y. & Yanai, K. *Geophys. Monogr. Ser.* **40**, 309–318 (1987).
- Boyd, F. R. & Gurney, J. J. *Science* **232**, 472–477 (1986).
- Gibbons, W. & Mann, A. *Geology* **11**, 3–6 (1983).
- Caby, R., Bertland, J. M. L. & Black, R. in *Precambrian Plate Tectonics* (ed. Kröner, A.) 407–434 (Elsevier, Amsterdam, 1981).
- Dobretsov, N. L. & Sobolev, N. V. *Ofioliti* **9**, 401–424 (1984).
- Dobretsov, N. L., Coleman, R. G., Liou, J. G. & Maruyama, S. *Ofioliti* **12**, 445–456 (1988).
- Dong, S., Shen, Q., Sun, D. & Lu, L. *Metamorphic Map of China, 1:4,000,000* (Geology Publishing House, Beijing, 1986).
- Ye, H. *Acta petrol. mineral. Anal.* **6**, 103–111 (1987).
- Liou, J. G., Wang, X., Coleman, R. G. & Maruyama, S. *Tectonics* **8**, 609–619 (1989).
- Xiong, J. & Wang, W. *Xinjiang Geol.* **4**, 33–46 (1986).
- Liou, J. G. *et al. Geology* **17**, 1127–1131 (1989).
- Nakajima, T., Banno, S. & Suzuki, T. *J. Petrol.* **18**, 263–284 (1977).
- Brown, E. H. *Geol. Soc. Am. Bull.* **85**, 333–344 (1974).
- Liou, J. G. & Maruyama, S. *J. metamorph. Geol.* **5**, 371–395 (1987).
- Jäger, E. in *Lectures in Isotope Geology* (eds Jäger, E. & Hunziker, J.) 13–26 (Springer, Berlin, 1979).
- Dodson, M. H. & McClelland-Brown, E. *Geol. Soc. Lond. Mem.* **10**, 315–325 (1984).
- Hongzhen, W. in *The Geology of China* (eds Yang, Z., Cheng, Y. & Wang, H.) 50–63 (Clarendon, Oxford, 1986).
- Cowie, J. W. & Johnson, M. R. W. *Geol. Soc. Lond. Mem.* **10**, 47–64 (1984).
- Ernst, W. G. *Geology* **16**, 1081–1084 (1988).
- Steiger, R. H. & Jäger, E. *Earth planet. Sci. Lett.* **36**, 359–362 (1977).

ACKNOWLEDGEMENTS. T.N. and S.U. thank K. Shibata, H. Aoyama and A. Miyamoto for advice and encouragement, help with mineral separation, and preparation of thin sections, respectively. This work was supported in part by the NSF.

Energetics of running: a new perspective

Rodger Kram & C. Richard Taylor*

Harvard University, Museum of Comparative Zoology,
Concord Field Station, Old Causeway Road, Bedford,
Massachusetts 01730, USA

THE amount of energy used to run a mile is nearly the same whether it is run at top speed or at a leisurely pace (although it is used more rapidly at the higher speed). This puzzling independence of energy cost and speed is found generally among running animals¹, although, on a per gram basis, cost is much higher for smaller animals. Running involves little work against the environment²; work is done by muscles and tendons to lift and accelerate the body and limbs. Some of the work is recovered from muscle-tendon springs without metabolic cost^{3,4} and work rate does not parallel metabolic rate with either speed or size. Regardless of the amount of work muscles do, they must be activated and develop force to support the weight of the body. Load-carrying experiments have shown that the cost of supporting an extra newton of load is the same as the weight-specific cost of running⁵. Size differences in cost are proportional to stride frequency at equivalent speeds, suggesting that the time available for developing force is important in determining cost^{6,7}. We report a simple inverse relationship between the rate of energy used for running and the time the foot applies force to the ground during each stride. These results support the hypothesis⁸ that it is primarily the cost of supporting the animal's weight and the time course of generating this force that determines the cost of running.

In testing the above hypothesis across both speed and size, we made three assumptions. First, we assumed that most of the

force exerted by the muscles acts to oppose gravity. This seems reasonable as force platform measurements on a variety of running animals have shown that the vertical force is at least 10 times the horizontal force⁹, and the average vertical force during a stride must be equal to the weight of the animal (W_b). Next, we assumed that a unit volume of active muscle exerts the same force on the ground regardless of speed or animal size. A recent study¹⁰ supports this assumption for both speed and size. Although the mechanical advantage is worse in smaller animals, they have correspondingly shorter muscle fibres¹¹, and force per cross sectional area is constant. Finally, we assumed that muscles operate over similar ranges of the force-velocity relationship, irrespective of speed and size. This requires faster fibres at higher speeds and in smaller animals, because the force is applied in a shorter period of time. It costs more for faster fibres to generate force because their cross-bridges cycle and consume ATP at faster rates^{12,13}.

The hypothesis states that the rate of energy consumption per newton of body weight by the muscles of a running animal ($\dot{E}_{\text{metab}}/W_b$) is inversely proportional to the weight-specific rate of force application, W_b/t_c divided by W_b , where t_c is the time for which the foot applies force to the ground during each stride. So

$$\dot{E}_{\text{metab}}/W_b = c \cdot 1/t_c \quad (1)$$

where c is a cost coefficient. This hypothesis can be tested by measuring $\dot{E}_{\text{metab}}$ and t_c over a range of speeds and gaits in animals of different sizes.

Our hypothesis should also account for the size-related difference in the cost of transport, the amount of energy consumed to move a unit weight a unit distance. The cost of transport, E_{trans}/W_b , is simply weight-specific rate of energy

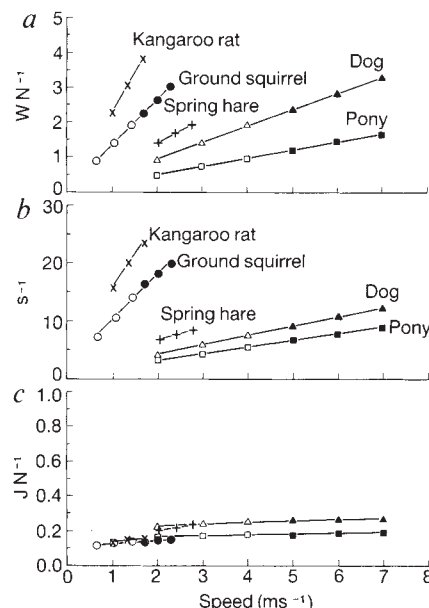


FIG. 1 Weight-specific rates of energy consumption ($\dot{E}_{\text{metab}}/W_b$, panel a) and force application ($1/t_c$, panel b) increase linearly over the entire range of running speeds in a diverse assortment of mammals, ranging in size from 30-g kangaroo rats to 140-kg ponies. The ratio of these two is a cost coefficient (panel c) that can be used to calculate the rate of energy consumption of a running animal knowing only body weight (W_b) and the time per stride that a single foot is in contact with the ground (t_c). The cost coefficient is nearly constant across speed and size, indicating that metabolic rate is inversely proportional to the time of force application. Open symbols, trotting; closed symbols, galloping; crossed symbols, bipedal hopping. The three points for each gait show the average for all the individuals (see text) at slow, medium and fast speeds within each gait. Lines are linear least-squares regressions.

* To whom correspondence should be addressed.

consumption divided by running speed, v :

$$E_{\text{trans}}/W_b = \frac{\dot{E}_{\text{metab}}}{W_b \cdot v} \quad (2)$$

Running speed is equal to step length, the distance travelled while each foot is in contact with the ground (L_c), divided by t_c :

$$v = L_c/t_c \quad (3)$$

We can combine equations (1), (2) and (3) to solve for E_{trans}/W_b in terms of L_c :

$$E_{\text{trans}}/W_b = c \cdot 1/L_c \quad (4)$$

Thus, we predict larger animals with longer legs and step lengths will have lower transport costs.

We used a comparative approach to test this hypothesis, taking advantage of the two- to three-fold change in metabolic rate with speed and a 10-fold difference in cost of transport across animal size. Steady-state oxygen consumption and average time of foot contact were measured over the range of aerobic running speeds in kangaroo rats (*Dipodomys merriami*, 32 g), ground squirrels (*Spermophilus tridecemlineatus*, 210 g), spring hares (*Pedetes capensis*, 3.0 kg), dogs (*Canis familiaris*, 25.8 kg) and ponies (*Equus caballus*, 141 kg). Oxygen consumption was measured while animals ran at a constant speed on a treadmill using an open flow system described by Fedak *et al.*¹⁴. We subtracted the extrapolated oxygen consumption at zero speed from the measured value, assuming that the additional energy is used by the muscles for running¹⁵. To calculate rate of energy consumption we assumed 20.1 joules of energy are liberated for each ml of oxygen consumed. The time of contact, t_c , was measured for each limb from high-speed ciné films (Photosonics 16 mm 1PL, 200 frames s⁻¹), or force platforms^{9,16} (in the case of the bipeds) and averaged for the total number of limbs. Oxygen consumption and t_c were measured on the same individuals. Some of the measurements of oxygen consumption have been reported previously^{17,18}. Weight-specific rates of energy consumption (Fig. 1a) and force application (Fig. 1b) both increased linearly with speed. Our measurements of oxygen consumption agree well with the general allometric equation¹⁹. Using equation (1), we calculated the cost coefficient for each animal at each speed, by dividing $\dot{E}_{\text{metab}}/W_b$ by $1/t_c$ (Fig. 1c). As predicted by the hypothesis, this coefficient was nearly constant across speed and about the same value for the different species.

We found that L_c increased only slightly over each animal's aerobic speed range (Fig. 2). This explains the initial paradox—why the same amount of energy is consumed in running a mile

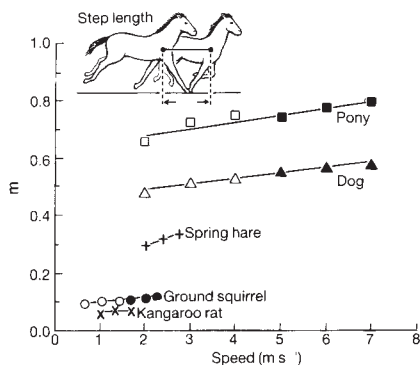


FIG. 2 Step length, L_c (the distance the body moves forward during the time a single foot is on the ground per stride), increases very little as speed increases but is greater in larger animals. Symbols here are as in Fig. 1. We use the term step length, as defined by Gray²², whereas others have used it to mean half of a stride.

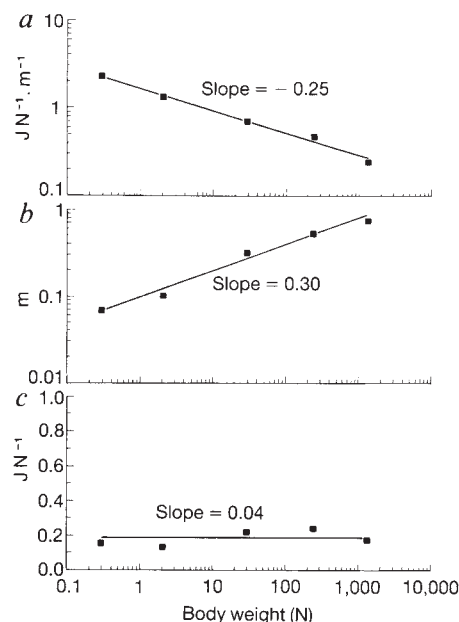


FIG. 3 Cost of transport (E_{trans}/W_b , panel a)—the amount of energy consumed in moving a unit of body weight a unit distance—decreases and step length (L_c , panel b) increases with increasing body weight. The product of cost of transport and step length is a cost coefficient (c , panel c) that is independent of animal size, indicating that cost of transport is inversely proportional to step length. The step-length value plotted in panel b is either the middle trotting or hopping speed.

regardless of speed. Using equations (2) and (3) we see that E_{trans}/W_b will be independent of speed if L_c is constant. Step length increased in a regular, almost geometrical manner with body weight. When plotted on logarithmic coordinates, L_c increases linearly with body weight with a slope of 0.30 (95% confidence limits, ± 0.05) (Fig. 3b). This slope agrees with values reported for other species²⁰. The cost of transport, E_{trans}/W_b , decreased with increasing body weight (Fig. 3a). When plotted on logarithmic coordinates this is a linear relationship with a slope of $-0.25 (\pm 0.09)$, a value similar to previous reports^{19,21}. The ratio of E_{trans}/W_b to $1/L_c$ equals c . As predicted by equation (4), c is nearly constant over the range of body size with no indication of a size dependency (mean of 0.183 J N^{-1} , s.d. of 0.045) (Fig. 3c). There is remarkably little variation in this cost coefficient, considering E_{trans}/W_b varies by 10-fold.

Apart from shedding light on the energetics of locomotion, our findings may be of practical use to ecologists interested in field estimates of energy consumption. A measurement of t_c from films and a knowledge of the weight of an animal can provide a good approximation of its rate of energy consumption and cost of transport. Our measurements of the metabolic cost of running and t_c show that the simple hypothesis formalized in equation (1) works remarkably well in mammals ranging in size from 30 g kangaroo rats to 140 kg horses. We conclude that the cost of running is primarily determined by the cost of supporting weight and by the time course of force application. □

Received 1 May; accepted 24 May 1990.

1. Full, R. J. in *Energy Transformations in Cells and Animals* (eds Wieser, W. & Gnaiger, E.) 175–182 (Thieme, Stuttgart, 1989).
2. Pugh, L. G. C. E. *J. Physiol., Lond.* **213**, 255–276 (1971).
3. Alexander, R. McN. *Am. Zool.* **24**, 85–94 (1984).
4. Blewener, A., Alexander, R. McN. & Heglund, N. C. *J. Zool.* **195**, 369–383 (1981).
5. Taylor, C. R., Heglund, N. C., McMahon, T. A. & Looney, T. R. *J. exp. Biol.* **86**, 9–18 (1980).
6. Heglund, N. C., Fedak, M. A., Taylor, C. R. & Cavagna, G. A. *J. exp. Biol.* **97**, 57–66 (1982).
7. Heglund, N. C. & Taylor, C. R. *J. exp. Biol.* **138**, 301–318 (1988).
8. Taylor, C. R. *J. exp. Biol.* **115**, 253–262 (1985).
9. Cavagna, G. A., Heglund, N. C. & Taylor, C. R. *Am. J. Physiol.* **233**, R243–R261 (1977).
10. Blewener, A. A. *Science* **245**, 45–48 (1989).

11. Alexander, R. McN., Jayes, A. S., Maloiy, G. M. O. & Wathutu, E. M. *J. Zool.* **194**, 539-552 (1981).
12. Huxley, A. F. *J. Physiol.* **243**, 1-43 (1974).
13. Rall, J. *Ex. Sports Sci. Rev.* **13**, 33-74 (1985).
14. Fedak, M. A., Rome, L. & Seeherman, H. J. *J. appl. Physiol.* **51**, 772-776 (1981).
15. Schmidt-Nielsen, K. *Science* **177**, 222-227, 1972.
16. Heglund, N. C. *J. exp. Biol.* **93**, 333-338 (1981).
17. Thompson, S. D., MacMillen, R. E., Burke, E. M. & Taylor, C. R. *Nature* **287**, 223-224 (1980).
18. Fedak, M. A. & Seeherman, H. J. *Nature* **282**, 713-716 (1979).
19. Taylor, C. R., Heglund, N. C. & Maloiy, G. M. O. *J. exp. Biol.* **97**, 1-21 (1982).
20. Alexander, R. McN., Langman, V. A., & Jayes, A. S. *J. Zool.* **183**, 291-300 (1977).
21. Taylor, C. R., Schmidt-Nielsen, K., & Raab, J. L. *Am. J. Physiol.* **219**, 1104-1107 (1970).
22. Gray, J. *Animal Locomotion* (Weidenfeld & Nicolson, London, 1968).

ACKNOWLEDGEMENTS. This study was supported by the US NIH.

Spatial representation of words in the brain implied by studies of a unilateral neglect patient

Alfonso Caramazza & Argye E. Hillis*

Cognitive Science Center, The Johns Hopkins University, Baltimore, Maryland 21218, USA

*HealthSouth Rehabilitation Corporation, 9512 Harford Road, Baltimore, Maryland 21234, USA

READING and writing require access to stored knowledge about the spelling of words. Presumably, we recognize *chair* but not *chare* or *chiar* as a word of English, and similarly would write 'chair' but not 'chare' or 'chiar', because we access orthographic representations that specify the identity and the order of the graphemes (abstract letter representations) that comprise the spelling of words^{1,2}. Thus, a fundamental problem concerns the content and structure of the hypothesized orthographic representations, and how information about grapheme order is represented and processed. We present evidence from a brain-damaged patient (N.G.) with unilateral neglect that this information is coded spatially. Unilateral neglect is a disorder clinically characterized by the inability to perceive or respond to stimuli presented to the side contralateral to the site of lesion, despite the absence of significant sensory or motor deficits^{3,4}. The patient made reading and spelling errors only on the right half of words, regardless of length. Furthermore, she produced the same pattern of errors in reading and spelling, irrespective of the topographic arrangement of stimuli in reading (horizontal, vertical or mirror-reversed words) and of the type of response in spelling (written, oral or backward oral spelling). This pattern of performance suggests that order information in orthographic representations is coded spatially in a word-centred coordinate system; that is, in a spatially defined coordinate frame whose centre corresponds to the midpoint of a canonical, orientation-invariant representation of the word and not the midpoint of the word stimulus⁵.

N.G. is a 77-year-old left-handed woman, who presented with a particularly pure clinical picture of unilateral neglect due to a stroke involving a large area of the left parietal white matter and the left anterior basal ganglia, adjacent to the head of the caudate nucleus (as revealed by CT scan). In contrast to her normal spoken language, eye movements and visual fields, N.G. showed severe difficulties in processing the right side of objects and words: in bisecting lines she systematically displaced the centre to the left; in drawing objects from memory she omitted or distorted details on the right side; and in reading and writing she made errors only on the right side of words. The right-sided reading errors persisted even when it could be demonstrated that she had processed the whole word, as revealed by her flawless ability to name all the letters in the word. Similarly, her bisection displacement persisted even after she successfully indicated the extremities of the line before bisecting it (see Fig. 1). Thus, N.G.'s unilateral neglect seemed to reflect a deficit in processing the right side of internal representations and not a deficit in sensory processing or stimulus scanning⁶⁻¹³.

N.G. was asked to read and write several thousand words of various lengths in various tasks ($n = 2,202$ and 1,662, respectively). The stimuli used in the various tasks were selected from word lists controlled for standard lexical and non-lexical factors known to affect reading and spelling performance, such as word and stem frequency, grammatical class, length, concreteness, and morphological structure. The lists were compiled from The Johns Hopkins Dyslexia and Morphology Batteries. Overall accuracy was 77% in reading and 24% in spelling. The only factor to affect performance in reading was frequency—N.G. read high-frequency words with greater accuracy than low-frequency words. The only factor to affect spelling was length, as she spelled short words with greater accuracy than long words. Furthermore, errors were only made on the right half of a word, and these increased linearly as a function of absolute distance in graphemes from the centre of the word (Table 1). This result was independent of the form of stimulus input in reading and the form of output in spelling, that is, N.G. produced virtually identical rates and distributions of errors in reading words presented in horizontal, vertical or mirror reversed form. She also produced essentially identical rates and distributions of errors in written, oral and backward-oral spelling of words. The distribution of errors as a function of letter position in a word was virtually identical across all reading and spelling tasks (Table 2). Examples of N.G.'s performance in the various reading and spelling tasks are shown in Fig. 1. In all tasks, her errors were restricted to the 'right' end of canonical representations of words.

Although other brain-damaged subjects have been described whose reading performance is similar to that of N.G.^{6,9,10,12,13}, none of these showed a spatially-specific deficit which (where tested) remained invariant—always involved the same end of the word—under topographic transformations of the stimulus

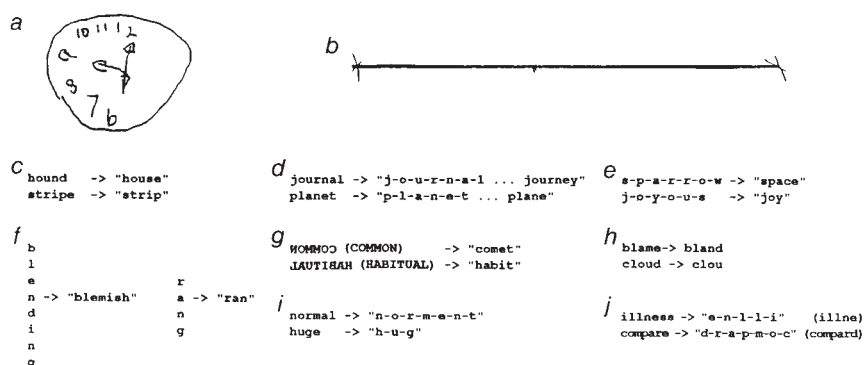


FIG. 1 Examples of N.G.'s performance in various tasks. *a* Attempt to draw a clock; *b* bisection of a 5-inch line after having indicated the extremities. The centre is shifted to the left, indicating neglect of the right end of the line; *c*, reading horizontally displayed words; *d*, reading horizontally displayed words after naming the letters in the words; *e*, naming orally spelled words; *f*, reading vertically displayed words; *g*, reading mirror-reversed words; *h*, written spelling; *i*, oral spelling; *j*, backward oral spelling. Note that in all cases errors only involved the right half of a canonical, word-centred representation of the word, and not the right half of a viewer-centred representation (compare horizontal and vertical, and mirror-reversed conditions in reading) or the last letters produced (compare oral and backward oral spelling).

TABLE 1 Distribution of errors as a function of letter position in words of different lengths

Reading										Spelling							
Position in word										Position in word							
	1	2	3	4	5	6	7	8	9		1	2	3	4	5	6	7
Word length										Word length							
4 (141)	0	0	4	15						4 (48)	0	2	13	25			
5 (219)	0	0	1	8	18					5 (70)	0	0	6	20	29		
6 (204)	0	0	0	4	8	25				6 (62)	0	0	5	15	26	39	
7 (82)	0	0	1	4	5	16	31			7 (67)	0	0	3	5	15	28	51
8 (88)	0	0	0	1	3	18	21	34									
9 (55)	0	0	0	0	2	10	23	28	37								

Reading										Spelling								
Word Centre										Word Centre								
	x											x						
4				0	0	4	15			4			0	2	13	25		
5			0	0	1	8	18			5		0	0	6	20	29		
6			0	0	0	4	8	25		6	0	0	0	5	15	26	39	
7			0	0	1	4	5	16	31	7	0	0	3	5	15	28	51	
8		0	0	0	1	3	18	21	34									
9	0	0	0	0	2	10	23	28	37									

The top part of each panel reports percentage errors at each letter position for 4- to 9-letter words in reading and 4- to 7-letter words in writing to dictation. The number of stimuli for each word length is shown in parentheses. The bottom of each panel reports the same data, with each word length arranged by reference to the centre of words. When performance is displayed in the latter fashion, it is apparent that reading and spelling errors occur almost exclusively on the right half of words, and increase linearly as a function of absolute distance in number of graphemes from the centre of the word.

(horizontal, vertical and mirror-reversed orientation). Thus, N.G. read the horizontally presented word *hound* as 'house' and the mirror-reversed word *common* as 'comet'. These errors involve the ends of words, despite the fact that their relative positions in the stimuli are on the right and left part of the stimuli, respectively. This pattern of performance contrasts with that of similar patients previously described. For example, patient V.B.¹² made errors on the left part of word stimuli—the beginnings of words—when they were presented in normal, horizontal orientation, and continued to make errors on the left part of word stimuli—the ends of words—when they were presented in mirror-reversed form. These errors involve the same relative position in the stimulus word and not the same position within a canonical, orientation-invariant representation of the word, as with N.G.

The contrasting results obtained for N.G. and V.B. and other patients with different forms of neglect dyslexia, provide evidence for a dissociation between different levels of representation in visual word recognition, that is between a canonical, word-centred representation and a stimulus- or viewer-centred representation of word stimuli⁵. N.G.'s performance contrasts with previously reported cases in that her spelling performance was qualitatively identical to her performance in reading, and was invariant for different forms of output (written or oral spelling, forward or backward spelling). These results severely constrain hypotheses about the nature of the deficit responsible for N.G.'s reading and spelling performance:

(1) The fact that N.G.'s reading impairment took the same form irrespective of topographic transformations of the stimulus input (such as, normal compared with mirror-reversed presentation), rules out the possibility that the deficit involves levels of representation where information is arrayed retinotopically or in a viewer-centred coordinate system¹². Instead, this fact implies that the deficit involves a level of processing where graphemic information is represented in an orientation-invariant canonical form, that is, in a word-centred coordinate system.

(2) The fact that her spelling deficit took the same form irrespective of whether a written or oral response was required, and that her performance remained invariant regardless of the order of output (forward or backward spelling)⁹, rules out a deficit to a serial order processing mechanism as well as a deficit

to premotor mechanisms for planning actions in space. Oral spelling does not involve planning actions in external space, yet errors were restricted to the right half of words.

(3) Because N.G.'s processing difficulties took the same form in reading and spelling, irrespective of the stimulus input or form of response, the deficit seems to concern a level of representation that has abstracted away from the specifics of the two sets of tasks. The deficit may concern a level of processing that specifies the identity and order of abstract letter representations and not one that encodes specific letter shapes.

(4) The points in (1) and the fact that N.G.'s reading and spelling processing deficit was restricted to the right end of words regardless of the form of input or output suggests that her difficulty involves processing the right part of a canonical orthographic representation in a word-centred coordinate system.

Facts (1)–(4), and the fact that reading performance remained invariant for normal and mirror-reversed words, and that reading and spelling performance were qualitatively identical, suggest that reading and spelling involve computing a spatially coded,

TABLE 2 Distribution of errors as a function of letter position

Task	Letter position in 6-letter words (% of total errors)					
	1	2	3	4	5	6
Regular reading	0	0	0	9	27	64
Vertical reading	0	0	2	10	35	52
Naming of orally spelled words	0	0	7	20	28	45
Mirror-reversed reading	0	1	9	16	29	46
Written spelling	0	0	6	19	31	46
Oral spelling	0	1	2	7	29	60
Backward spelling	0	0	1	12	34	53
Delayed copying	0	1	1	13	28	57

Percentage errors are reported for each letter position for the same set of 6-letter words ($n=108$) used in all reading and spelling tasks. The results show that errors are virtually restricted to the right half of each word, independently of the form of input in reading (horizontal, vertical or mirror-reversed written words or orally spelled words) and independently of the type (written or oral) and order (forward or backward) of output in spelling.

TABLE 3 Percentage correct stem responses

Stimulus type	n	% Correctly read stem
Unaffixed stem	96	75%
Legally suffixed stem ($\chi^2(1)=15.10, P<0.001$)	96	96%
Unaffixed stem	287	66%
Illegally suffixed stem ($\chi^2(1)=61.23, P<0.001$)	287	93%

Percentage correct stem responses for 'legally' suffixed words, 'illegally' suffixed words and the same stems without suffixes. Responses were scored as correct if the entire stem was produced (whether or not the suffix was read correctly). It was hypothesized that NG would correctly read the stem more often when a suffix was added, because a greater number of letters would be on the left half of the representation. For example, only 'lau' occurs on the left half of launch (which N.G. read as 'laundry'), where 'laun' (and half of c) occur on the left half of the illegally suffixed form, 'launchify' (which she read as 'launcher'). Generally, N.G.'s responses retained at least the left half of the stimulus: for example, she read 'child' as 'chill' but read its legally suffixed form, 'childishly' as 'childless'.

abstract orthographic representation in a word-centred coordinate system. For N.G., it would be expected that adding letters on the right end of a word should facilitate its reading because the centre of the internal representation of the longer graphemic string would be shifted rightward with the consequence that a greater part of the word falls in the left, normally processed half of the representation. For example, N.G. should be much more likely to produce the stem 'contrast' (with or without a suffix) when shown *contrastiveness* than when shown *contrast*.

She might read *contrastiveness* as 'contrasting' or 'contrastive', but in reading *contrast* she should produce responses such as 'continue' or 'control'. In several experiments, this expectation was fully supported, as she made significantly fewer reading errors on the stems of suffixed words (e.g. *harassment*) than unaffixed words (e.g., *harass*) (Table 3).

The task-independent but highly specific form of reading and writing deficit shows that one form of unilateral neglect results from a selective deficit in processing the right (or left) half of abstract, orientation-invariant, internal representations. This indicates that the order of graphemes in an orthographic representation is coded by reference to relative spatial positions in a word-centred coordinate system. □

Received 4 April; accepted 18 May 1990.

1. Caramazza, A. & Miceli, G. in *Brain and Reading* (eds von Euler, V. et al. 257-268 (Stockton Press, 1989).
2. Prinzmetal, W. et al. *J. Mem. Lang.* **25**, 461-475 (1986).
3. De Renzi, E. *Disorders of Space Exploration and Cognition* (Wiley, Chichester, 1982).
4. Heilman, K. M. et al. in *Clinical Neuropsychology* (2nd ed.) (eds Heilman, K. M. & Valenstein, E.) 243-293 (Oxford University Press, New York, 1985).
5. Monk, A. F. Q. *J. exp. Psychol.* **A37**, 613-625 (1985).
6. Kinsbourne, M. & Warrington, E. K. *J. Neurol. Neurosurg. Psychiatr.* **25**, 334-339 (1962).
7. De Renzi, E. et al. *Cortex* **6**, 191-203 (1970).
8. Bisiach, E. et al. *Brain* **102**, 609-618 (1979).
9. Baxter, D. & Warrington, E. K. *J. Neurol. Neurosurg. Psychiatr.* **46**, 1073-1078 (1983).
10. Costello, A. & Warrington, E. K. *J. Neurol. Neurosurg. Psychiatr.* **50**, 1110-1116 (1987).
11. Riddoch, M. J. & Humphreys, G. W. in *Neurophysiological and Neuropsychological Aspects of Spatial Neglect* (ed. Jeannerod, M.) 151-182 (North-Holland, Amsterdam, 1987).
12. Ellis, A. et al. *Cognitive Neuropsychol.* **4**, 439-464 (1987).
13. Behrmann, M. et al. *Brain* (in the press).

ACKNOWLEDGEMENTS. We thank N.G. for her participation in the study, and M. McCloskey, B. Rapp and S. Yantis for helpful comments on the paper. This work was supported by grants from the NIH, the Seaver Institute and the McDonnell-Pew Program in cognitive neuroscience.

Suppression by glutamate of cGMP-activated conductance in retinal bipolar cells

Scott Nawy & Craig E. Jahr

Vollum Institute, L474, Oregon Health Sciences University, Portland, Oregon 97201, USA

DEPOLARIZING bipolar cells (DBC) of the retina are the only neurons in the vertebrate central nervous system known to be hyperpolarized by the neurotransmitter glutamate¹. Both glutamate and its analogue L-2-amino-4-phosphonobutyrate (APB) hyperpolarize DBCs by decreasing membrane conductance²⁻⁵. Furthermore, glutamate responses in DBCs slowly decrease during whole-cell recording, suggesting that the response involves a second messenger system⁶. Here we report that intracellular cyclic GMP or GTP activates a membrane conductance that is suppressed by APB, resulting in an enhanced APB response. In the presence of GTP- γ -S, APB causes an irreversible suppression of the conductance. Inhibitors of G-protein activation or phosphodiesterase activity decrease the APB response. Thus, the DBC glutamate receptor seems to close ion channels by increasing the rate of cGMP hydrolysis by a G protein-mediated process that is strikingly similar to light transduction in photoreceptors.

Bipolar cells in retinal slices from larval tiger salamander were identified by the morphology of their cell bodies and their position in the inner nuclear layer. When voltage-clamped with patch pipettes at negative holding potentials, about half of the bipolar cells responded to glutamate and APB, an agonist known to act selectively on DBCs², with an outward current (Fig. 1a). Two other glutamate analogues, kainate (50 μ M) and quisqualate (5 μ M), did not evoke any response in these cells at concentrations which produced large inward currents in other retinal cell types (data not shown). Thus, cells responding to

APB were classified as DBCs. The outward current produced by APB at negative holding potentials is due to the reduction of a tonically active conductance with a reversal potential near 0 mV^{3,4}. The amplitude of the glutamate and APB response decreased with time during whole-cell recording⁶, even when the intracellular solution contained ATP and Mg²⁺, which prevent wash-out of M current in sympathetic ganglion neurons⁷ and GABA_A responses in hippocampal neurons⁸. The washout of the glutamate response coincided with a decrease in holding current and membrane conductance, suggesting that the standing inward current blocked by glutamate was no longer active.

To investigate the possibility that a second messenger is necessary to maintain the standing inward current, we tested the ability of intracellular cGMP to support the APB-sensitive conductance. In photoreceptors, cGMP is the second messenger which opens light-sensitive channels (reviewed in ref. 9). In DBCs dialysed with 1 mM cGMP a gradually increasing inward current that reached a maximal amplitude of 79.1 ± 11.3 pA (mean $\pm$ s.e.; 17 out of 20 DBCs) was observed (Fig. 1b). This inward current, as well as the standing inward current seen in control neurons, was suppressed by APB (Fig. 1b). Although the APB response did not completely block the cGMP-induced current in all cells, the amplitude of the APB response increased in parallel with the cGMP-induced current, from a mean initial value of $31.7 \text{ pA} \pm 5.0$ to a maximum of $71.6 \text{ pA} \pm 10.8$ ($n=17$). Dialysis with cGMP did not induce an inward current in cells that did not respond to APB, and presumably were not DBCs. The effect of APB on DBCs is remarkably similar to the interaction of light and cGMP in photoreceptors¹⁰⁻¹³.

Dialysis of DBCs with 5 mM GTP produced an inward current similar to that produced by cGMP (Fig. 1c). In all nine DBCs dialysed with GTP, the mean inward current reached a maximal amplitude of 99.6 ± 17.6 pA, whereas the APB response increased from 25.7 ± 6.8 pA to 50.3 ± 11.5 pA. The ability of high concentrations of GTP to mimic the actions of cGMP suggests that these cells contain an endogenous guanylate cyclase activity. Although it is possible that the inward current is

due to the stimulation of a G protein by GTP, this seems unlikely as the current was not regularly observed in cells that were dialysed with 100 μ M GTP, a concentration that is sufficient to maintain agonist responses in systems that require only G-protein function¹⁴.

Current-voltage (I - V) relations in a cGMP-dialysed cell measured immediately after break-in and following development of the inward current, revealed that the inward current was associated with a conductance increase (Fig. 1*d*). The I - V relation in the presence of APB was essentially the same as the I - V relation obtained before the inward current appeared (Fig.

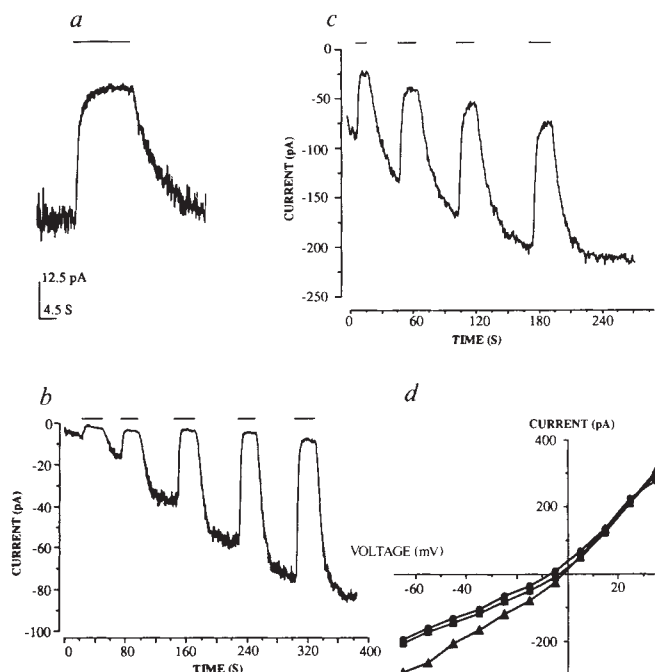


FIG. 1 Intracellular cGMP and GTP produce an inward current in DBCs. *a*, Response of a DBC to application of 1 μ M APB (bar). Cells were voltage-clamped at -65 mV (*a*-*d*). *b*, Intracellular dialysis of a DBC with 1 mM cGMP resulted in a slowly developing inward current that was reversibly blocked by 1 μ M APB (applied at the bars). Whole-cell recording established at $t=0$ (*b*, *c*). *c*, Dialysis of a DBC with 5 mM GTP also induced the slow development of an APB-sensitive inward current. The ordinate indicates total holding current (*c*, *d*). *d*, I - V plots from another DBC (dialysed with 1 mM cGMP), just after gaining access to the whole cell and before the inward current developed ($\bullet$), following development of inward current ($\blacktriangle$) and during application of 1 μ M APB ($\blacksquare$). In each condition, a family of voltage steps depolarized the neuron in 10-mV increments from a holding potential of -65 mV. The apparent changes in membrane noise associated with APB application (*a* and *b*) were not analysed, but suggest that the underlying channel conductance may be larger than that in photoreceptors. In control conditions, input resistance of DBCs ranged from 300–800 M Ω .

METHODS. Retinal slices from tiger salamander were prepared as described²³. Slices were cut to a thickness of 100 μ m and viewed on a Zeiss Standard microscope with a $\times 40$ water-immersion objective equipped with Hoffman optics. Patch-clamp recordings were made using a List EPC-7 amplifier and stored on video tape for later analysis. Pipette access to the cell's interior was continually monitored by recording the current responses to 10 mV $\times$ 10 ms voltage jumps. The bathing solution contained (in mM): NaCl, 108; MgCl₂, 1.2; KCl, 2.5; CaCl₂, 2.0; glucose, 5.0; HEPES buffer, 5.0, adjusted to pH 7.7 with NaOH and bubbled with 100% O₂. Picrotoxin (100 μ M), strychnine (1 μ M) and CoCl₂ (1 mM) were also added to block synaptic transmission in the slice. The pipette solution contained (in mM): Tris phosphate, 72; Tris base, 18; EGTA, 10; NaCl, 15; MgATP, 2; GTP, 0.1 and was adjusted with CsOH (95 mM) to a pH of 7.7. Pipette solutions containing 5 mM GTP required the addition of 10 mM CsOH to adjust the pH to 7.7. Drugs were applied to cells through a series of flow-pipes⁶. All chemicals were from Sigma except APB (Tocris Neuramin and Cambridge Research Biochemicals).

1*d*), suggesting that APB decreases the same conductance that is activated by cGMP. Similar results were obtained in five cGMP- and two GTP-dialysed cells.

The APB response seems to be mediated by a G protein because dialysis with the non-hydrolysable GTP analogue, GTP- γ -S, made the APB response irreversible. After formation of the whole cell recording with a pipette containing GTP- γ -S (100 μ M) and cGMP (1 mM) an inward current developed, as was typical of cGMP-loaded cells (Fig. 2*a*). Application of APB blocked the inward current as before. The inward current, however, was not restored on removal of APB, and further applications failed to produce any additional response. We included cGMP in the pipette with GTP- γ -S to enhance the APB response, which was small in these experiments, perhaps owing to the presence of extracellular glutamate released by the retinal tissue. These data suggest that APB closes channels by activating a G protein. In the presence of GTP- γ -S, this activation is irreversible. A similar result has been observed in photoreceptors where the introduction of GTP- γ -S into rod outer segments blocks recovery of the light response¹⁵.

Further support for the role of a G protein came from experiments with GDP- β -S, which blocks activation of G proteins in a competitive manner¹⁶. Binding of GDP- β -S to the G protein would thus be expected to prevent APB from closing channels. Figure 2*b* shows a cell that was dialysed with 500 μ M GDP- β -S and 1 mM cGMP. The first application of APB did not reverse the development of the inward current as it did in cells dialysed only with cGMP. The APB response was completely blocked within 2 min after formation of the whole-cell recording. These results were observed in all five cells in which GTP, which competes with GDP- β -S for the G protein-binding site, was excluded from the recording solution. GDP- β -S reduces the sensitivity of dialysed rod outer segments to light¹⁵, also con-

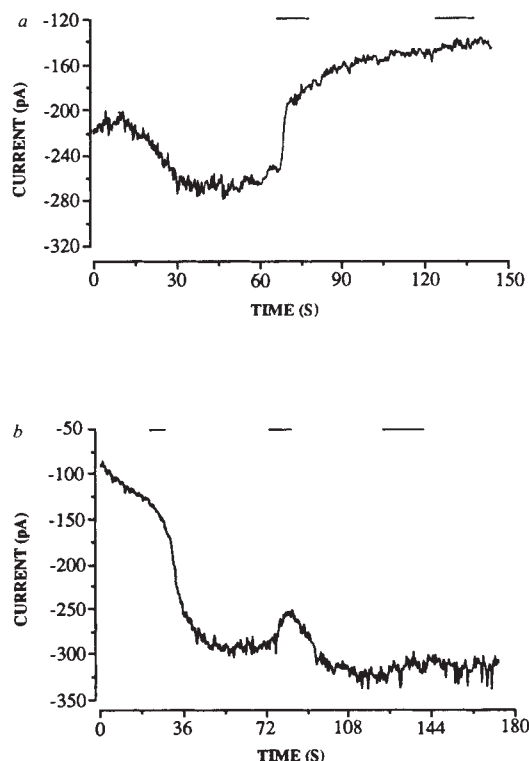


FIG. 2 Inhibition of the inward current by APB is mediated by a G protein. *a*, In a DBC dialysed with 100 μ M GTP- γ -S and 1 mM cGMP, the first application of 1 μ M APB (bars) irreversibly blocked the inward current. A second application of APB had no effect ($n=5$). *b*, The response of a different DBC dialysed with 500 μ M GDP- β -S and 1 mM cGMP to 1 μ M APB (bars). Holding potential for both cells was -65 mV.

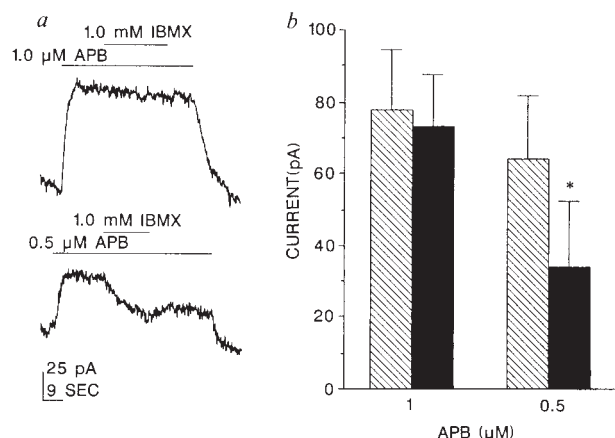


FIG. 3 IBMX partially blocks the response to submaximal concentrations of APB. *a*, Records from one cell comparing the effect of IBMX at 1.0 μ M APB (top) and 0.5 μ M APB (bottom). The pipette contained 1 mM cGMP in order to enhance the size of the APB response. Holding potential was -65 mV. *b*, Mean amplitude (four cells) of the response to 1 μ M (left) and 0.5 μ M (right) APB in the absence (hatched) and presence (dark bars) of 1 mM IBMX. IBMX reduced the response to 0.5 μ M APB by 48%. The standard errors reflect the variability of the APB responses from cell to cell, but within each cell, the reduction of the response to 0.5 μ M APB by IBMX was significant (*) ($P < 0.03$, paired *t*-test). Effects of IBMX on responses to both concentrations of APB were obtained in two cells.

sistent with a mechanism in DBCs similar to that of light transduction. Our results also suggest that the inward current activated by cGMP does not require a G protein, as it persisted in the presence of GDP- β -S.

In photoreceptors, activation of the G protein transducin results in the increased rate of hydrolysis of cGMP by a phosphodiesterase (reviewed in ref. 17). The phosphodiesterase inhibitor 3-isobutylmethylxanthine (IBMX) was used to test whether an APB-activated G protein might activate phosphodiesterase in DBCs. Figure 3*a* contrasts the effects of 1 mM IBMX on responses to 1 μ M (top) and 0.5 μ M (bottom) APB on the same cell. IBMX had little effect on the response to 1 μ M APB, but was able to inhibit substantially the response to 0.5 μ M APB. Figure 3*b* summarizes the effects of IBMX at two concentrations of APB. Overall, IBMX had no significant effect on responses to 1 μ M APB ($n = 4$), but reduced the response to 0.5 μ M by an average of 48% ($P < 0.03$, paired *t*-test; $n = 4$). The ability of IBMX to act only at lower concentrations of APB suggests that IBMX can block only a fraction of the phosphodiesterase activity in these cells, as in photoreceptors¹⁸. The effect of IBMX on APB responses is consistent with a role for phosphodiesterase, although we cannot rule out other possibilities such as antagonism of the APB receptor. These results suggest that phosphodiesterase can overcome the intracellular concentrations of cGMP attained during the course of these recordings.

In DBCs and photoreceptors, a tonically active inward current is shut off by glutamate and light, respectively. The essential steps in phototransduction have been clearly elucidated. Comparison of our results with those of previous electrophysiological studies of photoreceptors suggests that the two systems may share a number of properties. By substituting a glutamate receptor for rhodopsin, our data become consistent with the basic steps of the transduction cascade: activation of a G protein, which in turn activates a cGMP phosphodiesterase, resulting in the hydrolysis of cGMP and the closure of ionic channels gated by the nucleotide. The present model of phototransduction is based on a number of experimental approaches, including the demonstration that, in excised patches, cGMP can directly open channels that had properties similar to the light-sensitive channels¹⁹⁻²¹. Analogous experiments need to be performed on

DBC. The presence of a second messenger is also consistent with the longer delay in transmission from photoreceptors to DBCs than to horizontal or hyperpolarizing bipolar cells, the other cell types that receive direct synaptic input from photoreceptors²². □

Received 23 February; accepted 4 May 1990.

1. Murakami, M., Ohtsuka, T. & Shimazaki, H. *Vision Res.* **15**, 456-458 (1975).
2. Slaughter, M. M. & Miller, R. F. *Science* **211**, 182-185 (1981).
3. Shiels, R. A., Falk, G. & Naghshineh, S. *Nature* **294**, 592-594 (1981).
4. Nawy, S. & Copenhagen, D. R. *Nature* **325**, 56-58 (1987).
5. Attwell, D., Mobbs, P., Tessier-Lavigne, M. & Willson, M. *J. Physiol., Lond.* **387**, 125-161 (1987).
6. Nawy, S. & Jahr, C. E. *Neurosci. Lett.* **108**, 279-283 (1990).
7. Pfaffinger, P. J. *Neurosci.* **8**, 3343-3353 (1988).
8. Stelzer, A., Kay, A. R. & Wong, R. K. S. *Science* **241**, 339-341 (1988).
9. Yau, K. W. & Baylor, D. A. *Rev. Neurosci.* **12**, 289-327 (1989).
10. Yau, K. W. & Nakatani, K. *Nature* **317**, 252-255 (1985).
11. Cobbs, W. H., Barkdoll, A. E. & Pugh, E. N. *Nature* **317**, 64-66 (1985).
12. Matthews, H. R., Torre, V. & Lamb, T. D. *Nature* **313**, 582-585 (1985).
13. Kawamura, S. & Murakami, M. *J. gen. Physiol.* **94**, 649-668 (1989).
14. Trussell, L. O. & Jackson, M. B. *J. Neurosci.* **7**, 3306-3316 (1987).
15. Sather, W. A. & Detwiler, P. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9290-9294 (1987).
16. Eckstein, F., Cassel, D., Levkovitz, M., Lowe, M. & Selinger, Z. *J. Biol. Chem.* **254**, 9829-9834 (1979).
17. Hurley, J. B. *Rev. Physiol.* **49**, 793-812 (1987).
18. Ertel, E. A., Sather, W. A. & Detwiler, P. B. *Biophys. J.* **55**, 458a (1989).
19. Haynes, L. W., Kay, A. R. & Yau, K. W. *Nature* **321**, 66-70 (1986).
20. Zimmerman, A. L. & Baylor, D. A. *Nature* **321**, 70-72 (1986).
21. Matthews, G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 299-302 (1987).
22. Copenhagen, D. R., Ashmore, J. F. & Schnapf, J. K. *Vision Res.* **23**, 363-369 (1983).
23. Werblin, F. S. *J. Physiol., Lond.* **280**, 449-470 (1978).

ACKNOWLEDGEMENTS. This work was supported by the National Eye Institute (S.N.) and the National Institute of Neurological Disorders and Stroke, the McKnight Endowment Fund for Neuroscience and the Medical Research Foundation of Oregon (C.E.J.).

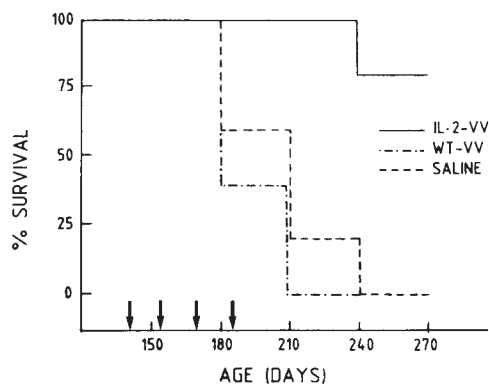
Recovery from autoimmunity of MRL/lpr mice after infection with an interleukin-2/vaccinia recombinant virus

Jose C. Gutierrez-Ramos, Jose L. Andreu, Yolanda Revilla, Eladio Viñuela & Carlos Martinez-A*

Centro de Biología Molecular del CSIC, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049-Madrid, Spain

INTERLEUKIN-2 (IL-2) is a T-cell derived molecule implicated in the clonal expansion of antigen-activated T cells¹ and in T-cell development². IL-2 is also implicated in autoimmune disease³, although its role is still controversial^{4,5}. Murine systemic lupus erythematosus (SLE) is a good model for human SLE as most of the immunological abnormalities in the human disease also seem to be operative in the mouse⁶. Among SLE mice, the MRL/lpr strain develops early in life autoimmune diseases such as immune complex-mediated glomerulonephritis, arthritis and arteritis^{7,8}. Lymphoid abnormalities associated with those diseases in this strain are thymic atrophy and abnormal proliferation of CD3⁺ CD4⁺ CD8⁻ 'double-negative' T cells, resulting in massive generalized lymph node enlargement. We have therefore now examined the effects of IL-2 on the disease progression in MRL/lpr mice⁹ using live vaccinia recombinant viruses expressing the human IL-2 gene¹⁰. Vaccinated mice showed prolonged survival, decreased autoantibody and rheumatoid factor titres, marked attenuation of kidney interstitial infiltration and intraglomerular proliferation, as well as clearance of synovial mononuclear infiltrates. Inoculation with the IL-2/vaccinia recombinant virus led, in addition, to drastic reduction of the double-negative T-cell population, improved thymic differentiation and restoration of normal values of mature cells in peripheral lymphoid organs.

* To whom correspondence should be addressed.



MRL/lpr (lymphoproliferation) mice infected with IL-2/vaccinia survived for more than a year (394 days) compared with MRL/lpr mice treated with saline or infected with wild-type vaccinia (mean survival time 195 days) (Fig. 1). Survival was accompanied by a lack of clinical symptoms as the arthritis, cutaneous ulcers and general deterioration present in normal lpr mice were barely detectable after treatment. The raised protein levels found in urine of non-infected MRL/lpr mice and MRL/lpr mice infected with wild-type vaccinia virus (300–500 mg per 100 ml), were significantly decreased and even normalized in the recombinant-treated group (<30 mg per 100 ml). The effect of IL-2/vaccinia infection on the levels of auto-antibodies was determined as previously described¹¹. This treatment had a clear effect on serum levels of IgM antibodies against double-stranded (ds) DNA, which were decreased from an absorbance of 0.444 ± 0.081 in the group infected with wild-type

FIG. 1 Mortality curves of IL-2/vaccinia (IL-2-VV)-infected mice compared with control groups. Three different groups of 4-month-old female MRL/lpr received four consecutive doses of wild-type vaccinia virus (WT-VV) or saline solution (arrows).

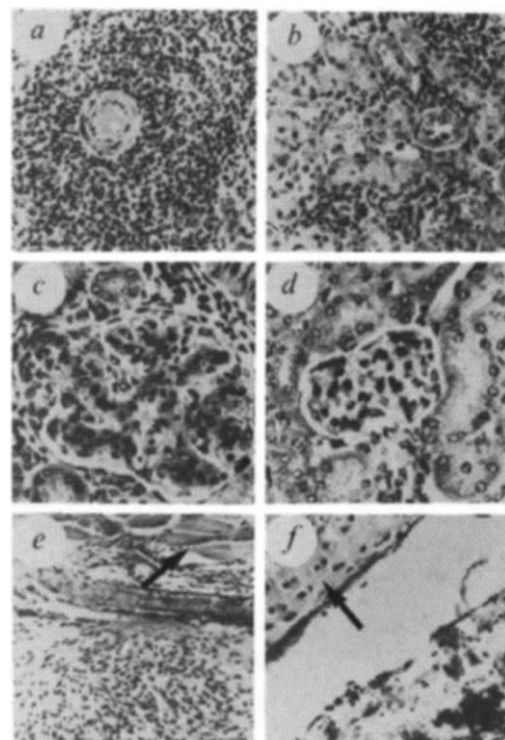
METHODS. Every 14 days, 10^7 plaque-forming units (PFUs) of recombinant IL-2/vaccinia virus (vCF13) in 0.4 ml saline solution were injected. Control age-matched females were treated in the same conditions with either 10^7 PFUs of wild-type vaccinia virus (strain WR) or 0.4 ml saline solution. The vCF13 virus is a recombinant vaccinia containing the human IL-2 gene²¹ under the control of the vaccinia p7.5 promoter¹⁰, isolated after CV-1 cell infection. Approximately 30 U ml^{-1} IL-2 were detected in supernatants of CV-1 cultures as measured by proliferation of murine IL-2-dependent T lymphocytes (CTLL-2). IL-2 bioactivity was detectable in the serum of these mice 10 days after intraperitoneal injection of vCF13, but not in those infected with wild-type vaccinia. In total, 67 animals in three independent experiments were studied. Mice were either infected with the IL-2/vaccinia recombinant virus (32 mice) or the wild-type vaccinia virus (25 animals), or injected with PBS (21 mice) and housed within adjacent cages in the same room, under sterile conditions provided by laminar flow hoods. Virus stocks were purified, characterized and titred on CV-1 cells as described previously²².

vaccinia virus to 0.138 ± 0.039 in the recombinant-treated mice. No differences in anti-dsDNA IgG were observed (absorbance 0.909 ± 0.218 versus 1.041 ± 0.224 for mice infected with wild-type and recombinant vaccinia virus, respectively). Also, there was a 75% reduction in the quantity of IgM rheumatoid factor which combines with IgG2a ($69.6 \pm 8.2 \mu\text{g ml}^{-1}$ versus $17.7 \pm 9.2 \mu\text{g ml}^{-1}$ for the recombinant- and wild-type-treated groups respectively), whereas no significant reduction in IgM rheumatoid factor anti-IgG1 was detected (20.3 ± 1.3 versus $21.6 \pm 2.3 \mu\text{g ml}^{-1}$).

Amelioration of the autoimmune disease was further confirmed by a minimal mononuclear infiltrate in kidney in the recombinant-treated mice (Fig. 2b) when compared with the massive interstitial infiltration characteristic of the MRL/lpr kidney, which was also observed in the wild-type-infected group (Fig. 2a). The increase in glomerular size and cellularity

FIG. 2 Representative photomicrographs of kidney (interstitium and glomerulus) and synovium from IL-2/vaccinia (IL-2-VV) and wild-type vaccinia (WT-VV)-infected mice. *a*, Massive mononuclear interstitial infiltration in a kidney section from an lpr mouse infected with wild-type vaccinia virus. *b*, Kidney interstitium from a IL-2/vaccinia-treated mouse (magnification, $\times 300$). Glomeruli from animals infected with wild-type vaccinia or recombinant vaccinia are shown in *c* and *d*, respectively (magnification, $\times 500$). The mononuclear infiltration present in the synovial membrane of recombinant-infected mice is shown in *e*. *f*, Representative interface between synovium and cartilage in a recombinant-infected mouse (magnification, $\times 300$). Portions of muscle and cartilage (arrows) are included in the pictures to locate the synovium.

METHODS. Kidney and synovial samples were fixed in Bouin's solution and 10% (v/v) buffered formalin respectively, and routinely embedded in paraffin. Sections (4- μm thick) were prepared from each sample and stained with haematoxylin-eosin, Periodic acid-Schiff (PAS), Masson or Jones silver methenamine stain.



(mesangial and endothelial) characteristic of MRL/lpr was still present in sections from wild-type-infected mice (Fig. 2c), whereas the intraglomerular proliferation as well as glomerular enlargement were absent in kidneys from recombinant-infected animals (Fig. 2d). Immunofluorescence studies of kidney tissue sections revealed a lack of granular deposits of C3 and IgG in the glomeruli of recombinant-infected mice; however, they were detectable in the mesangium and capillary wall of control groups (not shown). Additionally, MRL/lpr mice develop swollen joints of the hind feet, pannus formation, proliferation of synovial tissue and degradation of articular cartilage⁶, none of which were present in the recombinant-treated animals. Synovial membrane infiltration studies revealed the persistence of a massive mononuclear infiltrate in wild-type-infected animals (2e), which was completely cleared in the joints from the recombinant-infected mice (Fig. 2f).

MRL/lpr disease is a result of a defect in pre-T stem-cell differentiation requiring thymic influence for the development

of the full-blown syndrome. This syndrome is associated with the expansion of phenotypically immature double-negative cells as well as a marked decrease of immature 'double positive' thymocytes¹². We have now explored the effects induced by the recombinant infection on thymic differentiation. Two-colour flow-cytometric analysis of thymocytes showing correlated expression of CD4 and CD8 is depicted in Fig. 3a. Thymuses from recombinant-infected mice showed a significant reduction in double-negative cells (9 versus 28%) accompanied by a three-fold increase in the immature double-positive CD4⁺CD8⁺ population (16 versus 44%). Mature CD3⁺ cells (Fig. 3a, b) were also affected. Among them, the CD4⁺CD8⁻ subset, which is moderately expanded in thymuses from lpr-bearing SLE mice¹³, was reduced (44 versus 29%), the CD4⁻CD8⁺ subset remaining unaffected. The enlargement in the CD4⁺CD8⁺ population observed in Fig. 3a is further confirmed in Fig. 3b, where CD4 and CD8 positive cells as well as CD3⁻ or CD3^{duil} cells show an increase from 12% to 34%. Figure 3c also shows that the thymus from recombinant-infected mice contained up to 30% of Thy1⁺CD3⁻ cells, which constituted a minute population in wild-type-infected mice. These data support the assertion that the IL-2/vaccinia recombinant treatment promotes the intrathymic differentiation-maturation of early T-cell precursors, resulting in a threefold expansion of Thy1⁺CD3⁻CD4⁺CD8⁺ cells.

As previously described⁶, control mice showed hyperplasia of peripheral lymphoid organs produced by the accumulation of lymphocytes of phenotype CD3⁺CD4⁻CD8⁻ (Fig. 4). These cells are not actively cycling¹⁴ and they seem to migrate and accumulate in these organs. In fact, treatment of mice with

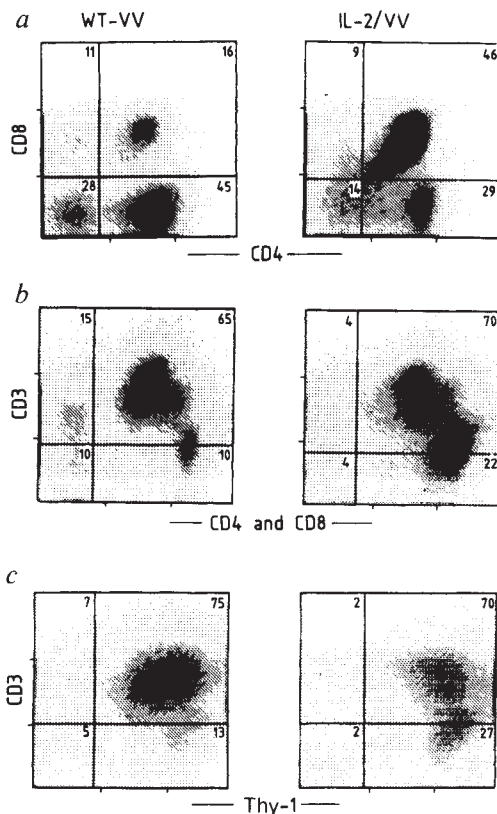


FIG. 3 Phenotype of T cells from thymuses of lpr mice infected with wild-type vaccinia (WT-VV) or IL-2/vaccinia (IL-2/VV). Thymuses were obtained from 5-month-old MRL/lpr mice infected with wild-type vaccinia or 9-month-old mice infected with vCF13 recombinant vaccinia. Cells (52×10^6) were obtained for the wild-type-infected mice and 360×10^6 cells were obtained for the recombinant-infected mice. Thymus cells (10^6) were stained with: a, phycoerythrin (PE) conjugated anti-CD4 (Becton Dickinson) and fluorescein isothiocyanate (FITC) conjugated anti-CD8 (B.D.); b, biotin-conjugated anti-CD8 (Becton Dickinson) followed by streptavidin-PE (Southern Biotechnology), PE-conjugated anti-CD4 and FITC-labelled anti-CD3 (2c11; ref. 23); or c, biotin-conjugated anti-Thy-1.2 (Becton Dickinson) followed by streptavidin-PE and FITC-labelled anti-CD3. Background values were obtained using irrelevant isotype-matched monoclonal antibodies and the respective second labelled antibodies or streptavidin-PE. Flow cytometric analysis was performed with an EPICS Profile Analyser (Coulter) after accumulation of 5×10^4 viable cells. Three-parameter data was obtained using logarithmic amplification of forward and side scatter. Each profile represents fluorescence data list mode files after gating-out dead cells. The figure shows the results of one typical experiment out of four.

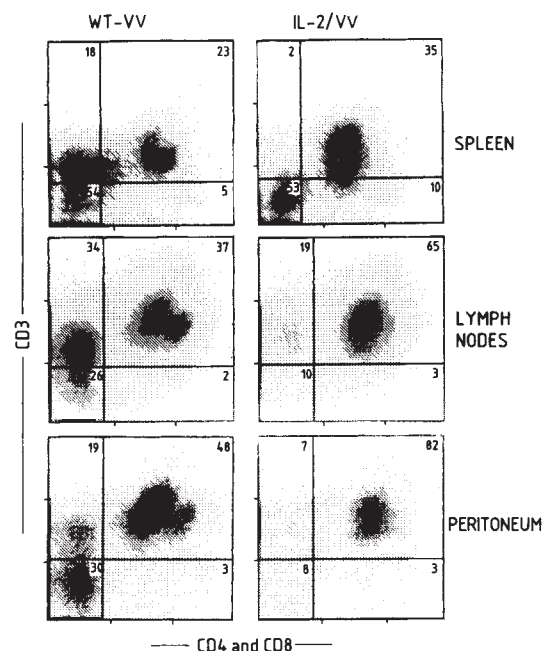


FIG. 4 Flow cytometric analysis showing the reduction of CD3⁺CD4⁻CD8⁻ T cells in spleen, lymph nodes and peritoneum of IL-2/vaccinia-treated mice. Cell suspensions from lymph nodes and spleen were obtained by standard procedures, and peritoneal cells were collected after extensive perfusion of the peritoneal cavity. The number of cells obtained in the wild-type- and in the recombinant-infected mice, respectively were 340×10^6 vs 150×10^6 for the spleen, 460×10^6 vs 240×10^6 for the inguinal lymph nodes, and 3×10^6 vs 5×10^6 for the peritoneum. Cells (5×10^5) for two-colour analysis were prepared in essentially the same way as for Fig. 3, using the same FITC-labelled anti-CD3, phycoerythrin-conjugated anti-CD4 and biotin-conjugated anti-CD8. Irrelevant isotype-matched mouse monoclonal antibodies were used in each experiment for negative controls. The data are representative of three different experiments.

anti-homing receptor antibody (MEL-14) results in reduction of lymph-node size¹². To explore the effect of recombinant infection on the peripheral accumulation of double-negative cells, spleen, lymph node, and peritoneal cells were stained with anti-CD3, anti-CD4 and anti-CD8 monoclonal antibodies. Figure 4 shows the drastic reduction of the CD3⁺CD4⁻CD8⁻ population in the recombinant-infected mice. The amount of CD3⁺ double-negative cells was reduced from 20% to 3% in the spleen, from 59% to 10% in lymph nodes, and from 17% to 4% in the peritoneum. This effect was accompanied by higher levels of mature cells (both CD4⁺CD8⁻ and CD4⁻CD8⁺) in all organs checked (Fig. 4), the CD4⁺CD8⁺ subset showing a clear reduction in the density of CD8 molecules on the cell surface. No double-positive T cells were found in the periphery when expression of CD4 versus CD8 was studied (not shown). Detailed phenotypic characterization of cells from these organs revealed that among the double-negative population, the CD3^{bright} and the CD3^{dull}/CD3⁻ cells had disappeared. Splenocytes stained for surface immunoglobulins showed that the B cells account for the CD3⁻ cells seen after treatment (not shown). We have not so far studied in depth the mechanisms by which recovery occurs, but it is certainly associated with an improvement in T-cell function and does not require expansion of natural killer (NK) cells (not shown). In agreement, all recombinant-infected mice resolved the pox lesions developed as a consequence of infection faster than their wild-type-infected counterparts. Recombinant IL-2/vaccinia infection has previously been shown to protect immunodeficient mice from the lethal effect of vaccinia virus, an effect not associated with antibodies^{10,15}.

The double-negative T cell population expanded in the lpr mice expresses $\alpha\beta$ T-cell receptors displaying autoreactive specificities^{16,17} related to the predominant (70%) expression of V β 8.2 and V β 8.3 gene products^{18,19}. IL-2/vaccinia inoculation resulted in an expansion of cells expressing the V β 8.1 gene and this correlated with a decrease in cells displaying the V β 8.2 and V β 8.3 specificities (not shown). These results reflect the absence of expansion of autoreactive clones that, together with the efficient abrogation of double-negative cells, makes the procedure described here extremely useful in those diseases in which there is selective expansion of cells with the double-negative phenotype. Human lupus erythematosus, where there is an up to 10-fold increase in such cells in peripheral blood, or rheumatoid arthritis, which shows an accumulation of such cells in the joints, are diseases easily envisaged as possible targets for studies in this direction, especially as IL-2/vaccinia infection prevents the development of kidney and synovial pathology. Furthermore, some of the anomalies in the immunopathogenesis of AIDS have been associated with deficient IL-2 production, as well as selective expansion of double-negative cells²⁰, characteristics also found in the experimental model studied here. □

21. Clark, F. et al. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2543-2547 (1984).
22. Chakrabarti, S., Brechling, R. & Moss, B. *Molec. cell. Biol.* **5**, 3403-3409 (1985).
23. Leo, O. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1374-1391 (1984).

ACKNOWLEDGEMENTS. We thank Drs B. Askonas and A. Mitchison for critical reading of the manuscript, Dr B. Moss for the gift of vCF13 recombinant vaccinia, Drs J. Kappler and P. Marrack for the gift of antibodies, and Drs R. Palacios, L. Pezzi, M. L. Toribio and D. Thorpe for help and comments. We thank E. Leonardo, M. J. Busto, J. A. Rodriguez and C. Martin for technical support, the Department of Molecular Biology of the Centro Nacional de Virología de Majadahonda for animal facilities and M. Messman for editorial assistance. This work was supported by Fondo de Investigaciones Sanitarias Comision Interministerial de Ciencia y Tecnología and the European Communities.

Mast cells as a source of both preformed and immunologically inducible TNF- α /cachectin

John R. Gordon & Stephen J. Galli*†

* Departments of Pathology, Beth Israel Hospital and Harvard Medical School, and the Division of Experimental Pathology, Beth Israel Hospital, Boston, Massachusetts 02215, USA

TUMOUR necrosis factor- α (TNF- α)/cachectin is a multifunctional cytokine that has effects in inflammation, sepsis, lipid and protein metabolism, haematopoiesis, angiogenesis and host resistance to parasites and malignancy¹⁻³. TNF- α was first described in activated macrophages¹⁻³, but certain mouse or rat mast cell populations (reviewed in refs 4, 5) and some *in vitro*-derived human cells with cytochemical features of mast cells—basophils⁶ may also contain products similar to TNF- α . Here we present evidence that resident mouse peritoneal mast cells constitutively contain large amounts of TNF- α bioactivity, whereas cultured, immature mast cells vary in their TNF- α content. IgE-dependent activation of cultured or peritoneal mast cells⁷ induces extracellular release of TNF- α and augments levels of TNF- α messenger RNA and bioactivity. These findings identify mouse mast cells as an important source of both preformed and immunologically inducible TNF- α , and suggest that release of TNF- α by mast cells may contribute to host defence, the pathophysiology of allergic diseases and other processes dependent on TNF- α .

The amount of TNF- α bioactivity expressed constitutively in different mouse mast cell populations varies (Table 1). Little or no activity was detected in interleukin-3 (IL-3)-dependent clones Cl-MC/9 or PT-18, whereas substantially higher levels were detected in an IL-3-independent clone (Cl-MC/C57.1) and in IL-3-dependent, bone marrow-derived cultured mast cells (BMCMC) from 4-6-week-old primary cultures. Levels of TNF- α bioactivity in unstimulated peritoneal mast cells greatly exceeded those in unstimulated cultured mast cells and were about twice those in lipopolysaccharide (LPS)-stimulated unfractionated peritoneal cells (Table 1) or LPS-stimulated Pu5-1.8 mouse macrophage-like cells ($1.74 \pm 0.07 \times 10^{-6}$ units per cell). Because some of the 'TNF- α -like' bioactivity of mast cells is associated with the cytoplasmic granules⁴, variation in mast cell TNF- α content may in part reflect the degree of maturation of the cells. IL-3-dependent mouse mast cells generally have an immature appearance by ultrastructure and contain much lower levels of cytoplasmic granule-associated mediators than do mature peritoneal mast cells⁸⁻¹⁰.

Stimulation of all three mast cell clones or BMCMC with IgE and antigen significantly ($P < 0.001$) increased levels of TNF- α bioactivity in cell lysates (Table 1). Virtually all the cytotoxicity could be inhibited by a rabbit anti-TNF antiserum (Fig. 1a and b) or by two other rabbit anti-TNF antisera (one provided by M. Palladino and one purchased from Genzyme, data not shown). Unfractionated peritoneal cells containing only 1-2%

Received 1 December 1989; accepted 26 April 1990.

1. Moller, G. (ed.) *Immun. Rev.* **63**, 1-209 (1981).
2. Toribio, M. L. et al. *Nature* **342**, 82-85 (1989).
3. Altman, A. et al. *J. exp. Med.* **154**, 791-808 (1981).
4. Linker-Israeli, M. et al. *J. Immun.* **130**, 2651-2655 (1983).
5. Weston, K. M. et al. *J. Immun.* **131**, 734-742 (1988).
6. Theofilopoulos, R. & Dixon, F. J. *Adv. Immun.* **37**, 269-390 (1985).
7. Katagiri, T., Cohen, P. L. & Eisenberg, R. A. *J. exp. Med.* **167**, 741-751 (1988).
8. Miescher, G. et al. *J. Immun.* **138**, 1959-1967 (1987).
9. Dixon, F. J. *Immun. Today* **2**, 145-147 (1981).
10. Flexner, C., Hugin, A. & Moss, B. *Nature* **330**, 259-261 (1987).
11. Katagiri, T. et al. *J. exp. Med.* **167**, 741-751 (1988).
12. Mountz, J. et al. *J. Immun.* **140**, 2943-2949 (1988).
13. Cohen, P. L., Rapoport, J. & Eisenberg, R. A. *Clin. Immun. Immunopath.* **40**, 485-491 (1986).
14. Budd, R. C. et al. *J. Immun.* **135**, 3704-3711 (1985).
15. Ramshaw, I. A. et al. *Nature* **329**, 545-546 (1987).
16. Wadsworth, S., Yui, K. & Greene, M. I. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1018-1022 (1988).
17. Sakaguchi, S. et al. *J. exp. Med.* **161**, 72-87 (1985).
18. Singer, P. A. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7018-7022 (1986).
19. Nemazee, D. A. et al. *Eur. J. Immun.* **15**, 760-764 (1988).
20. Marcos, M. A. R. et al. *Scand. J. Immun.* **25**, 321-333 (1987).

† To whom correspondence should be addressed.

mast cells expressed little or no cytotoxicity except after exposure to LPS (Fig. 1c), a treatment that stimulates peritoneal macrophages to synthesize TNF- α (refs 1-3). In three separate experiments, purified peritoneal mast cells (91-94% mast cells) expressed significant cytotoxicity, even when the mast cells were isolated in the presence of actinomycin D (which prevents induction of TNF- α mRNA), a treatment that blocked LPS-

induced TNF- α production by macrophages (Fig. 1c). Fractions of the same peritoneal cells depleted of mast cells (0.8% mast cells) expressed little or no cytotoxicity (Fig. 1c).

In each mast cell population tested, stimulation with IgE and antigen (as in Table 1) induced extracellular release of TNF- α bioactivity. The specific release of TNF- α bioactivity 60 min after antigen challenge (percentage release in antigen-stimulated cells minus percentage release in control unstimulated cells, mean $\pm$ s.e.m., $n=4-9$) exceeded baseline release during the same period by 2- to 5-fold and was $21 \pm 1\%$ for Cl-MC/C57.1, $35 \pm 2\%$ for PT-18, $27 \pm 5\%$ for BMCMC and $36 \pm 4\%$ for purified peritoneal mast cells. Moreover, the total amount of TNF- α bioactivity (total TNF- α) present in the cells and their supernatants 60 min after antigen challenge exceeded the corresponding values for the unstimulated cells (percentage increase in total TNF- α : Cl-MC/C57.1, +79%; PT-18, +28%, BMCMC, +118%; peritoneal mast cells, +62%). These results indicate that immunological stimulation induced both the release and synthesis of TNF- α . The antigen dose requirements for TNF- α and histamine release, as assessed with Cl-MC/C57.1 cells, were similar, although near-maximal levels of histamine or TNF- α release occurred at different intervals (5 min versus ≥ 30 min, respectively; J.R.G., S. Emery and S.J.G., manuscript in preparation).

TABLE 1 TNF- α /cachectin content of mouse mast cells from various sources

Cell type	Cell treatment	TNF- α content (units per 10^6 cells)
Cl-MC/C57.1	None	0.60 ± 0.06
	IgE-DNP-HSA	1.85 ± 0.02
Cl-MC/9	None	ND
	IgE-DNP-HSA	0.54 ± 0.04
PT-18	None	0.06 ± 0.04
	IgE-DNP-HSA	0.29 ± 0.08
BMCMC	None	0.53 ± 0.06
	IgE-DNP-HSA	1.89 ± 0.01
Total peritoneal cells (~60% macrophages, 1-2% mast cells)	None	ND
	LPS	1.74 ± 0.04
Peritoneal mast cells (91-94% mast cells, <1% macrophages)	LPS + act. D	ND
	act. D	2.98 ± 0.21

Bone marrow derived cultured mast cells (BMCMC) (>95% mast cells by neutral red staining) and mast cell lines^{4,8-10} were maintained, sensitized with IgE anti-DNP antibody²⁸ and stimulated for 4 h with DNP₃₀₋₄₀HSA (Sigma; 10 ng ml^{-1} cells) as described previously¹³. Unstimulated cells were treated identically, except for omission of IgE and antigen during incubations. Total peritoneal cells from intraperitoneal lavage of retired breeder BALB/c mice (Charles River Laboratories, Andover, Massachusetts) with Ca^{2+} /Mg²⁺-free Hanks' balanced salt solution (HBSS, Gibco) containing 1% FCS and 10 U ml^{-1} heparin (complete HBSS) were fractionated on 23% (w/v) metrizamide (Nycomed As, Norway) in complete HBSS into mast cell-enriched and mast cell-depleted fractions. In some experiments, actinomycin D (act.D, $10 \mu\text{g ml}^{-1}$) was included in lavage and density gradient media²⁹. Lipopolysaccharide (LPS) ($1 \mu\text{g ml}^{-1}$) stimulation was for 4 h. All cells were sonicated at 5×10^7 cells ml^{-1} in DMEM medium plus $1 \mu\text{g ml}^{-1}$ actinomycin D, then centrifuged at $2,000g$ for 15 min at 4°C . Supernatants were tested in quadruplicate for cytotoxic activity against L929 cells (ATCC, Maryland), at 6×10^4 target cells per well in 96-well plates, using an MTT assay^{4,30}. Recombinant mouse TNF- α (rmTNF, Genzyme, Boston) or cell lysates were incubated with L929 cells at 37°C for 18-24 h. 3-[4,5-dimethyl-thiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT; $500 \mu\text{g ml}^{-1}$) was added for 1-3 h, then levels of cell-incorporated MTT were measured with an automated ELISA plate reader. Results were expressed as per cent cytotoxicity, where per cent cytotoxicity = $(A_{570} \text{ control wells} - A_{570} \text{ experimental wells}) / A_{570} \text{ control wells} \times 100$. Cell lysates were tested in several dilutions, and their TNF- α content was calculated from values that gave <90% cytotoxicity. One unit of TNF- α was defined as that capable of causing 50% cytotoxicity. Internal controls of rmTNF- α were included in each experiment.

ND, not detectable.

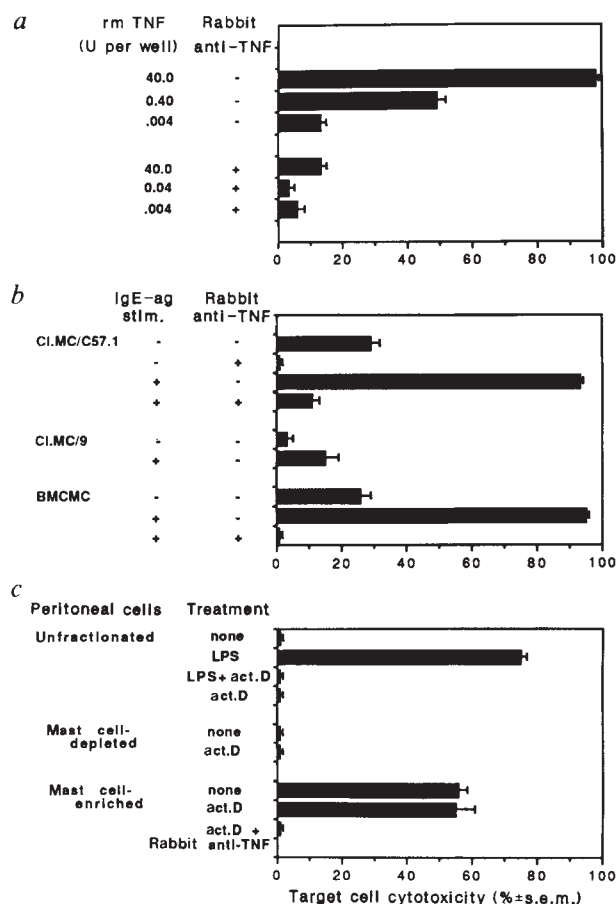


FIG. 1 *a*, Rabbit anti-recombinant murine (rm) TNF- α (rabbit anti-TNF) antiserum neutralized the cytotoxic activity of recombinant mouse TNF- α (rmTNF) against L929 cells, but pre-immune serum from the same rabbit did not (data not shown). *b*, Cell-associated TNF- α bioactivity is inducible by immunological activation of cultured mouse mast cells. Lysates (1×10^6 cell equivalents per well) from unstimulated or IgE- and antigen-stimulated (2 h) mast cells (IgE-ag stim.) were prepared and tested as in Table 1. Rabbit anti-TNF antiserum (1/20 final dilution) neutralized the cytotoxic activity. *c*, Freshly isolated, unstimulated peritoneal mast cells, but not peritoneal macrophages, constitutively contain high levels of TNF- α . The peritoneal lavages and cell fractionations were performed with medium $\pm$ actinomycin D²⁹, as in Table 1. Cell lysates were from total peritoneal cells (10^6 cell equivalents per well, 1-2% mast cells) incubated $\pm$ LPS ($1 \mu\text{g ml}^{-1}$) for 4 h, mast cell-depleted fractions of the density gradients (1.6×10^6 cell equivalents per well, 0.8% mast cells), and purified peritoneal mast cells (3.1×10^5 - 3.9×10^5 cell equivalents per well, 91-94% mast cells) tested $\pm$ rabbit anti-TNF antiserum (1/20 final dilution).

METHODS. Rabbit anti-TNF antiserum was prepared using rmTNF (Genentech) that had been reduced and alkylated with 50 mM DTT and 100 mM iodoacetate, respectively. New Zealand white rabbits were primed by intramuscular injection of 250 μg of protein in Freund's complete adjuvant and boosted with 50 μg protein in Freund's incomplete adjuvant. When titred against rmTNF in an L929 cell cytotoxicity assay, the rabbit anti-TNF antisera had an activity of $>8 \times 10^5$ neutralizing units ml^{-1} . The cells and cytotoxicity assay are described in Table 1. In *a* and *b*, the L929 assay culture volumes were 200 μl per well, in *c*, the volumes were 35 μl per well. Data shown are mean $\pm$ s.e.m. ($n=4-16$). Statistical analysis of the data were performed using the two-tailed Student's *t*-test.

Stimulation of mast cells by IgE and antigen markedly increased levels of TNF- α mRNA, compared with those detectable in unstimulated cells (Fig. 2a). Similar results were obtained when RNA from Cl-MC/C57.1 mast cells or Pu5-1.8 macrophages were probed with a 30-residue oligomer (30-mer) specific for mouse TNF- α but not TNF- β (lymphotoxin)¹¹ (Fig. 2c). Incubation of mast cells with either IgE (Fig. 2c, lane 2) or antigen alone (not shown) did not increase TNF- α mRNA levels above those in unstimulated cells. Very high levels of TNF- α mRNA were observed when immunological stimulation was done in the presence of cycloheximide (Fig. 2d, compare lanes 1 and 2) or when cells were stimulated with 12-*O*-tetradecanoylphorbol-13-acetate (TPA) (Fig. 2d, lane 3). In purified peritoneal mast cells incubated with cycloheximide, TNF- α mRNA was detectable in cells stimulated with either a monoclonal rat anti-mouse IgE or TPA (Fig. 2e, lanes 2 and 3) but not in unstimulated cells (Fig. 2e, lane 1). No TNF- α mRNA was detectable in mast cell-depleted, IgE-incubated fractions of peritoneal cells challenged with anti-IgE (data not shown). In Cl-MC/C57.1 mast cells, IgE-dependent stimulation of TNF- α mRNA was maximal at $\sim\frac{1}{2}$ to 1 h after stimulation with antigen, whereas highest levels of mRNA for IL-3 and IL-5 in the same cells were observed at ~ 2 h; the mRNA half-life for all three

cytokines, as determined following actinomycin D treatment of IgE- and antigen-stimulated cells was ~ 30 min (data not shown).

IgE-dependent stimulation of certain mouse mast cell lines^{12,13} or primary cultures of IL-3-dependent mouse mast cells^{13,14} increases levels of mRNA for several cytokines other than TNF- α and, in some cases, releases the products. The phenotype of mast cell populations *in vivo* can, however, differ significantly from that of mast cells maintained *in vitro* (reviewed in refs 9 and 10). Indeed, unlike the other cultured mast cells we tested, unstimulated PT-18 cells lacked detectable TNF- α mRNA (data not shown) and at least one long-term mast cell line did not contain detectable TNF- α mRNA even after stimulation¹². We find that freshly isolated peritoneal mast cells, as well as certain cultured mast cells, represent important sources of TNF- α bioactivity; and unlike resting macrophages¹⁻³, T cells¹⁵ or B cells¹⁶, which contain little or no TNF- α bioactivity, resident peritoneal mast cells store notable quantities of TNF- α available for immediate release.

TNF- α can promote local infiltration of inflammatory cells^{1-3,17-20}. Thus, IgE-dependent release of mast cell TNF- α may contribute to the leukocyte infiltration observed in immune responses to parasites²¹ or in late-phase reactions, such as those involved in asthma and other allergic diseases²². In support of such a role, we have shown that our rabbit anti-TNF antiserum diminishes IgE- and mast cell-dependent leukocyte infiltration into the skin of mice²³. The production of TNF- α by mast cells may also be important in other biological responses in which both TNF- α ^{1-3,24} and mast cells^{10,25,26} have been implicated, including host responses to tumours, regulation of wound healing and angiogenesis, and modulation of haematopoiesis and immunity. Nevertheless, we have not discounted the possibility that macrophage- and mast cell-derived TNF- α exhibit differences, such as those based on alternative mRNA splicing and/or glycosylation^{11,27}, that might result in different patterns of bioactivity. □

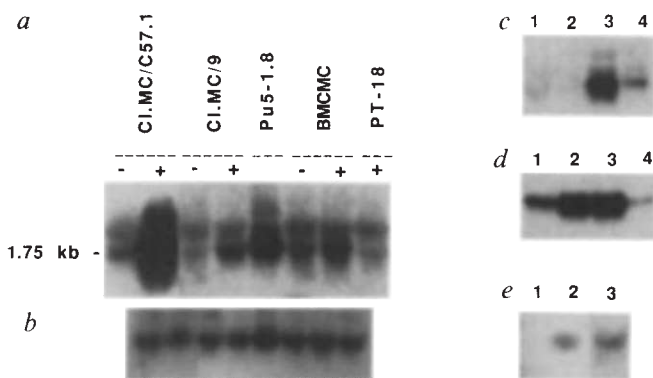


FIG. 2 Northern analyses of mast cell RNA using 1,000-base (a, d and e) or 30-base (c) cDNA probes specific for mouse TNF- α . a, RNA (20 μ g per lane) from mast cells tested after ± 4 h of stimulation with IgE, antigen and 3 μ g of oligo-dT-selected³¹ RNA from Pu5-1.8 macrophage-like cells stimulated for 4 h with TPA and A23187 (Sigma; 50 and 100 ng ml⁻¹, respectively). b, Same blot as in a, but stripped and probed for β -actin³². c, RNA (20 μ g per lane) from Cl-MC/C57.1 cells tested either without stimulation (lane 1), 4 h after incubation with IgE alone (lane 2), or 4 h after stimulation with IgE and antigen as in a (lane 3). Lane 4 contains 3 μ g of oligo-dT-selected RNA from Pu5-1.8 cells stimulated as in a. d, RNA (2.5 μ g per lane) from Cl-MC/C57.1 cells stimulated for 90 min with IgE and antigen in the absence (lane 1) or presence (lane 2) of cycloheximide (5 μ g ml⁻¹) or with TPA (50 ng ml⁻¹) (lane 3), or from unstimulated cells (lane 4). e, All of the available RNA (<1 μ g per lane) from $1.4\text{--}1.8 \times 10^6$ purified peritoneal mast cells (>86% purity) either not stimulated (lane 1) or stimulated for 5 h in the presence of cycloheximide (5 μ g ml⁻¹) either with a rat monoclonal anti-mouse IgE (150 ng ml⁻¹) after sensitization with IgE as in Table 1 (lane 2) or with TPA (50 ng ml⁻¹, lane 3).

METHODS. Mast cells were maintained and stimulated as in Table 1. Pu5-1.8 cells²⁷ (ATCC, Maryland) were maintained in RPMI 1640 medium containing 10% normal horse serum and 5×10^{-5} M 2-mercaptoethanol. Cellular RNA was prepared from cultured mast cells¹³ or peritoneal mast cells³³, as noted. Blots a, d and e were hybridized at 42 °C with a 1,000-base *Eco*RI murine TNF- α cDNA probe (from plasmid pMuTNF (ref. 27), Genentech) ³²P-labelled by random hexamer priming (specific activity $1\text{--}10 \times 10^6$ d.p.m. ng⁻¹ cDNA), washed 2×5 min at 20 °C in $2 \times$ SSC buffer, 0.1% SDS, for 2×30 min at 42 °C in $0.2 \times$ SSC buffer, 0.1% SDS, and exposed to Kodak X-OMAT-AR film. Blot c was hybridized at 62 °C with a ³²P-labelled, synthetic 30-mer (5'-GTCCCCCTCTCTCAGCTGGAAGACTCTCC-3', Pfizer, Groton, Connecticut) 3' labelled by the terminal deoxynucleotidyl transferase reaction (specific activity $\sim 10^6$ d.p.m. ng⁻¹ cDNA).

Received 17 July 1989; accepted 8 May 1990.

- Old, L. J. *Science* **230**, 630-632 (1985).
- Beutler, B. & Cerami, A. *Adv. Immun.* **42**, 213-231 (1988).
- Goeddel, D. V. *et al. Cold Spring Harb. Symp. quant. Biol.* **51**, 597-609 (1986).
- Young, J. D.-E., Liu, C.-C., Butler, G., Cohn, Z. A. & Galli, S. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9175-9179 (1987).
- Richards, A. L., Okuno, T., Takagaki, Y. & Djou, J. Y. *J. Immun.* **141**, 3061-3066 (1988).
- Steffen, M., Abboud, M., Potter, G. K., Yung, Y. P. & Moore, M. A. S. *Immunology* **66**, 445-450 (1989).
- Ishizaka, T. in *Allergy, Principles and Practice*, 3rd edn (eds Middleton, Jr. E., Reed, C. E., Ellis, E. F., Adkinson, Jr. N. F. & Yunginger, J. W.) 71-93 (Mosby, St Louis, 1988).
- Nabel, G., Galli, S. J., Dvorak, A. M., Dvorak, H. F. & Cantor, H. *Nature* **291**, 332-334 (1981).
- Galli, S. J. *et al. J. Cell Biol.* **95**, 435-444 (1982).
- Galli, S. J. *Lab. Invest.* **62**, 5-33 (1990).
- Semon, D., Kawashima, E., Jongeneel, C. V., Shakhov, A. N. & Nedospasov, S. A. *Nucleic Acids Res.* **15**, 9083-9084 (1987).
- Plaut, M. *et al. Nature* **339**, 64-67 (1989).
- Burd, P. *et al. J. exp. Med.* **170**, 245-257 (1989).
- Wodnar-Fillipowicz, A., Heusser, C. H. & Moroni, C. *Nature* **339**, 150-152 (1989).
- Cuturi, M. C. *et al. J. exp. Med.* **165**, 1581-1594 (1987).
- Sung, S.-S. J. *et al. J. exp. Med.* **168**, 1539-1551 (1988).
- Gamble, J. R., Harlan, J. M., Klebanoff, S. J. & Vadas, M. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8667-8671 (1985).
- Pober, J. S. *et al. J. Immun.* **137**, 1893-1896 (1986).
- Tracey, K. J. *et al. J. exp. Med.* **167**, 1211-1227 (1988).
- Rampart, M., DeSmet, W., Fiers, W. & Herman, A. G. *J. exp. Med.* **169**, 2227-2232 (1989).
- Askenase, P. W. *Springer Semin. Immunopathol.* **2**, 417-442 (1980).
- Lichtenstein, L. M. *J. Allergy clin. Immun.* **83**, 814-820 (1988).
- Galli, S. J., Wang, Z.-S., Gordon, J. R. & Wershi, B. K. *FASEB J.* **4**, A1875 (1990).
- Leibovich, S. J. *et al. Nature* **329**, 630-632 (1987).
- Selye, H. *The Mast Cells* (Butterworths, Washington, DC, 1965).
- Starkey, J. R., Crowle, P. K. & Taubenberger, S. *Int. J. Cancer* **42**, 48-52 (1988).
- Pennica, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6060-6064 (1985).
- Liu, F.-T. *et al. J. Immun.* **124**, 2728-2737 (1980).
- Johnston, P. A., Jansen, M. M., Somers, S. D., Adams, D. O. & Hamilton, T. A. *J. Immun.* **138**, 1551-1558 (1987).
- Green, L. M., Reade, J. L. & Ware, C. F. *J. immunol. Meth.* **70**, 257-268 (1984).
- Davis, L. G., Dibrner, M. D. & Battey, J. F. *Basic Methods in Molecular Biology* (Elsevier, New York, 1986).
- Cleveland, D. W. *et al. Cell* **20**, 95-105 (1980).
- Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156-159 (1987).

ACKNOWLEDGEMENTS. We thank T. Ishizaka, D. Pluznik and D. Rennick for gifts of cell lines, Fu-Tong Liu and D. Katz for the hybridoma H1-DNP-e-26, F. Finkelman for the monoclonal rat anti-mouse IgE, M. Palladino and Genentech Inc. for TNF- α cDNA, rmTNF- α , and rabbit anti-rmTNF- α antiserum and G. Andrews and Pfizer for the TNF- α 30-mer. This work was supported by the NIH.

HIV requires multiple gp120 molecules for CD4-mediated infection

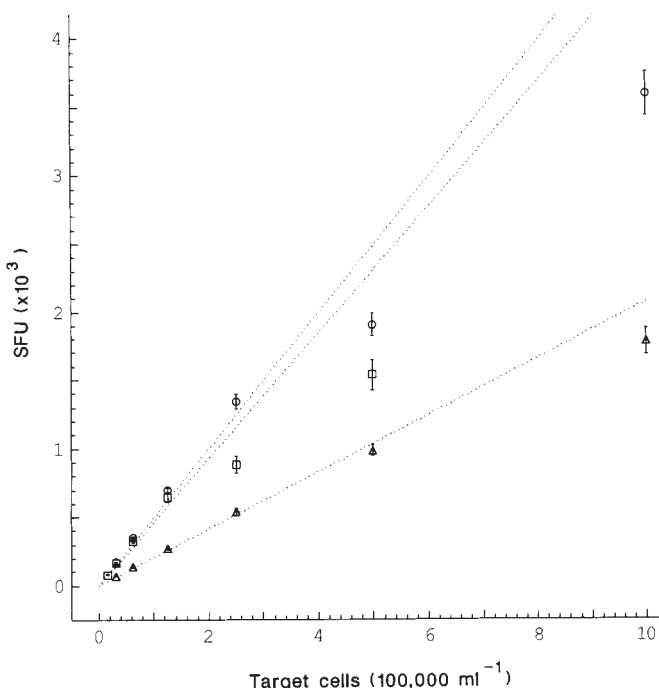
Scott P. Layne*, Michael J. Merges†, Micah Dembo*, John L. Spouge‡ & Peter L. Nara†

* Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

† Virus Biology Section, Laboratory of Tumor Cell Biology, NCI-FCRF, Frederick, Maryland 21701, USA

‡ National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland 20894, USA

BINDING of glycoprotein gp120 to the T cell-surface receptor CD4 is a crucial step in CD4-dependent infection of a target cell by the human immunodeficiency virus (HIV)¹⁻⁵. Blocking some or all gp120 molecules on the viral surface should therefore inhibit infection. Consequently, competitive receptor inhibitors, such as soluble synthetic CD4 (sCD4), synthetic CD4 peptides and immunoglobulins, have been investigated *in vitro*⁶⁻¹⁷ and *in vivo*¹⁸⁻²⁰, but little is known about the molecular mechanisms of these inhibitors. We have now quantitatively examined blocking by soluble CD4 in the hope of gaining insight into the complex process of viral binding, adsorption and penetration. At low sCD4 concentrations, the inhibition in three HIV strains is proportional to the binding of gp120. The biological association constant (gp120-sCD4 K_{assoc}) for HIV-2_{NIH2} is $(8.5 \pm 0.5) \times 10^7 \text{ M}^{-1}$, whereas K_{assoc} for HIV-1_{HXB3} (1.4 ± 0.2) and HIV-1_{MN} (1.7 ± 0.1) $\times 10^9 \text{ M}^{-1}$ are 15–20-fold larger. For all three viral strains, the biological K_{assoc} from infectivity assays is comparable to the chemical K_{assoc} . The inhibitory action of sCD4 at high concentrations, however, is not fully explained by simple proportionality with the binding to gp120. Positive synergy in blocking of infection occurs after about half the viral gp120s molecules are occupied, and is identical for all three viral strains, despite the large differences in K_{assoc} . Our method of measuring the viral-cell receptor K_{assoc} directly from infectivity assays is applicable to immunoglobulins, to other viruses and to assays using primary or transformed cell lines.



A quantitative infectivity assay requires the number of infectious events to be linear (unsaturated) in the target cell concentration^{21,22}. Figure 1 shows results for six concentrations of CEM-SS cells^{23,24}. Results were linear at lower cell concentrations and showed only minor assay saturation at higher concentrations. The multiplicity of infection in all assays was less than 0.025 (see Fig. 1, methods).

To distinguish among the possible mechanisms by which sCD4 blocks gp120-mediated infection, we previously analysed a kinetic model of the initial events of infection^{21,22}, in which each gp120 monomer within an oligomer is functionally independent and equivalent. This neutral hypothesis predicts that for an unsaturated assay, a plot of inverse infection ($1/\text{SFU}$, where SFU are syncytial-forming units) versus sCD4 concentration should be a straight line. That is, inhibition is proportional to binding. Dividing the slope of an inverse infection plot by its intercept gives the gp120-sCD4 K_{assoc} . Upward curvature in the inverse infection plot indicates positive synergy between gp120 molecules as they promote infection or between sCD4

FIG. 1 Examining the linearity of infectivity assays. To compare one assay with another, they must scale linearly with target cell concentration, because saturated assays make blockers seem ineffective^{21,22}. The graphs for HIV-1_{HXB3} (○), HIV-1_{MN} (□) and HIV-2_{NIH2} (△) show SFU plotted against target cell concentration. For each viral strain, assays were conducted at six target-cell and seven sCD4 concentrations (see Table 1). All results are the mean of eight microtitre wells (bars show ± 1 s.d.). Results without sCD4 are shown; similar results were obtained when sCD4 was added to the assays. The dotted lines are weighted least-squares fits. At all sCD4 concentrations, correlation coefficients were ≥ 0.89 for HIV-1_{HXB3}, ≥ 0.97 for HIV-1_{MN} and ≥ 0.98 for HIV-2_{NIH2}, substantiating that the infectivity assays were linear at the lower, more heavily weighted, cell concentrations.

METHODS. The HIV-1 and HIV-2 stocks were acutely collected from H9 cells to optimize infectivity²³. HIV-1 *in vivo* is a mixture of closely related viral subpopulations with minor envelope variations³¹. To address the influence of envelope microheterogeneity on infectivity³²⁻³⁴, we studied both a molecularly cloned viral stock (HIV-1_{HXB3}) and two uncloned viral stocks (HIV-1_{MN} and -2_{NIH2})²⁶. HIV-1_{MN} was selected because of its dominance in recent seroprevalence surveys³⁵. In typical viral stocks, gp120 is not only present on virions as oligomers³⁶ but also exists as soluble gp120 monomers³⁷. Thus, soluble gp120 may cause artefacts if it competes significantly with viral-associated gp120 for sCD4 or cell-surface CD4. In the present study, radioimmunoassay³⁸ was used to measure the total gp120 (soluble plus viral-associated) in all HIV stocks tested. The total gp120 in the viral inocula was 1×10^{-10} , 6×10^{-11} and $3 \times 10^{-12} \text{ M}$ for HIV-1_{HXB3}, HIV-1_{MN} and HIV-2_{NIH2}, respectively. These concentrations of total gp120 were, in the worst case, more than sevenfold smaller than $1/K_{\text{assoc}}$. Hence, the association of soluble gp120 with sCD4 or cell-surface CD4 was negligible in all viral stocks²³ used in this study. Infectious events were quantified by a modified version of the syncytium-forming assay^{23,24}. The modifications minimized artefacts associated with high cell concentrations in conjunction with sCD4 in the assay. Before the assay, CEM-SS cells were grown at low concentrations ($< 5 \times 10^5 \text{ ml}^{-1}$) to assure logarithmic growth. On the day of an assay, cells were suspended in fresh medium at densities of 4.5×10^5 and $2 \times 10^6 \text{ ml}^{-1}$ to serve as 'indicator' and 'target' cell stocks, respectively. Target cells (0.5 ml) were then transferred to culture tubes containing six different volumes of fresh medium and seven different sCD4 concentrations (total of 42 tubes). The final target-cell densities in the culture tubes were 1×10^6 , 5×10^5 , 2.5×10^5 , 1.25×10^5 , 6.25×10^4 and $3.13 \times 10^4 \text{ ml}^{-1}$. The respective reaction volumes were 1, 2, 4, 8, 16 and 32 ml. Identical multiplicities of infection (in graded volumes of viral stock) were added to each tube, resulting in a constant viral inoculum to reaction volume of 10%. To assure uniform mixing, culture tubes were rolled (~ 10 turns per min) during the 2-hour infection period. Next, the infected target cells were washed three times (centrifugation for 5 min, 200g, followed by suspension in 40 ml fresh medium and another centrifugation) and suspended in fresh medium at $5 \times 10^4 \text{ cells ml}^{-1}$. This thorough washing of sCD4 from the target cells prevented subsequent inhibition of syncytium formation in the monolayer of target and indicator cells (two washes removed all sCD4, data not shown). Cell monolayers were prepared by adding 4.5×10^4 indicator cells (0.1 ml) and 5×10^3 target cells (0.1 ml) to flat-bottomed microtitre wells (indicator:target cell ratio of 9:1). A total of eight wells were plated per sCD4 and target-cell concentration. Syncytia (representing the infection of individual target cells by cell-free virus) were counted on days three (HIV-2) or five (HIV-1) after infection. The above description applies to HIV-1_{HXB3} assays (an indicator to target-cell ratio of 90:1 gave identical results to the 9:1 ratio used here, data not shown). For HIV-1_{MN} assays, target-cell stocks were prepared at $1 \times 10^6 \text{ cells ml}^{-1}$ and target-cell densities in the culture tubes were twofold smaller. For HIV-2_{NIH2} assays, indicator-cell stocks were prepared at $6.5 \times 10^5 \text{ cells ml}^{-1}$; infected target-cells were washed and suspended at $1 \times 10^5 \text{ cells ml}^{-1}$ (indicator:target-cell ratio of 6.5:1). For assays with reaction volumes of 1, 2, 4, 8, 16 and 32 ml, the comparative number of SFU were calculated by multiplying actual number of SFU in each well by 32, 16, 8, 4, 2 and 1, respectively. Minimizing $f(a, b) = \sum \{ [y_i - (ax_i + b)] / \sigma_i \}^2$ gives a weighted least-squares linear fit to the data (y_i is the mean value of SFU at the i th sCD4 concentration, σ_i is the s.d. of SFU, and x_i is the blocker concentration at the i th data point). Terms in the correlation coefficient³⁹ were weighted by $1/\sigma_i^2$. A weighted correlation coefficient = 1 means that the least-squares line fits the data perfectly.

molecules as they bind to and block gp120. Downward curvature indicates corresponding negative synergy.

Figure 2 shows the inverse infection plot for HIV-2_{NIH2} at 1×10^6 target cells per ml. The plot curves progressively upward at higher sCD4 concentrations, indicating positive synergy in sCD4 blocking. Nevertheless, the data are linear at lower blocker concentrations and allow the determination of K_{assoc} . For example, the straight line in Fig. 2 gives a K_{assoc} of about $8.56 \times 10^7 \text{ M}^{-1}$ and the five other target cell concentrations gave comparable K_{assoc} (Table 1). K_{assoc} was also independent of the target cell concentration for the two strains of HIV-1 studied.

A chemical association constant for gp120-sCD4 binding cannot be presumed to measure a biological K_{assoc} , but it can be compared with the biological K_{assoc} as shown in Table 1. Despite a 20-fold difference in K_{assoc} between the viral types, the chemical measurements agree quite closely with our own. This difference agrees with reports of reduced blocking activity by sCD4 for HIV-2 in comparison with HIV-1^{12,25,41}. Thus, the initial slope of the inverse infection plot^{21,22} demonstrates that the biological activity of sCD4 (at low concentrations) is primarily due to reversible blocking of viral gp120 cell-surface CD4 interactions.

To analyse the synergistic blocking activity at higher concentrations, we normalized the inverse infection plot as follows. The inverse infection is divided by its control—inverse infection without blocker. The concentration of sCD4 is multiplied by

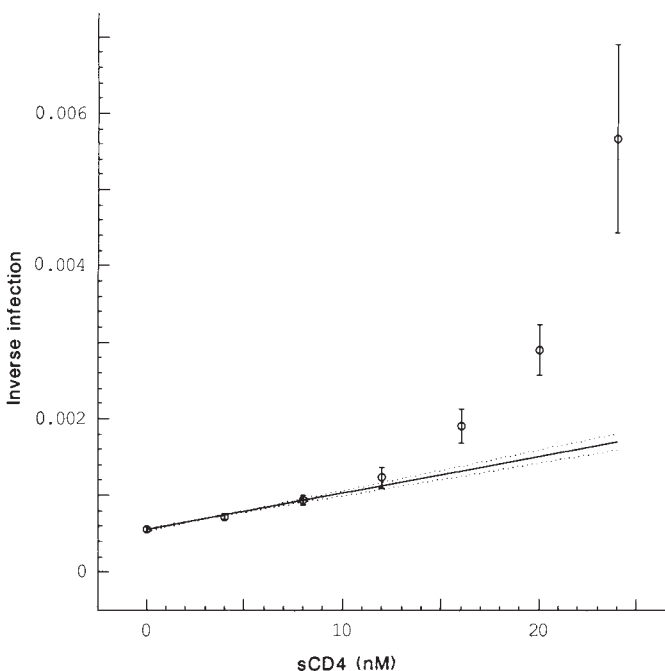


FIG. 2 Calculating the gp120-sCD4 K_{assoc} from infectivity assays. When infectivity assays scale linearly with target cell concentration, a plot of inverse infection ($1/\text{SFU}$) versus sCD4 concentration should yield a straight line with slope/intercept $\approx K_{\text{assoc}}$ (refs 21, 22). An example of this inverse infection plot is illustrated for HIV-2_{NIH2}, when target cell concentration is 1×10^6 cells ml^{-1} . The solid line is a weighted least-squares fit to the data (bars show mean ± 1 s.d.) at sCD4 concentrations of 0, 4, 8 and 12 nM. The fit has weighted correlation coefficient of ~ 0.99 , slope of $\sim 4.75 \times 10^4 \text{ M}^{-1}$, and intercept of $\sim 5.56 \times 10^4$, giving $K_{\text{assoc}} \approx 8.56 \times 10^7 \text{ M}^{-1}$ (y_i is the mean value of $1/\text{SFU}$, see Fig. 1 methods). The dotted lines are the 95% confidence limits for the fit. Both limits were calculated by a standard bootstrap method⁴⁰ and, in this example, give $7.47 \times 10^7 \leq K_{\text{assoc}} \leq 8.86 \times 10^7 \text{ M}^{-1}$. For the three viral strains, we calculated K_{assoc} from the lower four sCD4 concentrations in Table 1. The weighted correlation coefficients for all the fits are ≥ 0.85 for HIV-1_{HXB3}, ≥ 0.98 for HIV-1_{MN} and ≥ 0.99 for HIV-2_{NIH2}, indicating nearly complete explanations of the data at the lower sCD4 concentrations.

TABLE 1 Summary of the gp120-sCD4 K_{assoc}

Viral strain	Target cells (ml^{-1})	$K_{\text{assoc}} \pm 1 \text{ s.d. (M}^{-1}\text{)}$	Mean $K_{\text{assoc}} \pm 1 \text{ s.d. (M}^{-1}\text{)}$
HIV-1 _{HXB3}	3.13×10^4	$(1.38 \pm 0.17) \times 10^9$	$(1.4 \pm 0.20) \times 10^9$
	6.25×10^4	$(1.34 \pm 0.15) \times 10^9$	
	1.25×10^5	$(1.34 \pm 0.17) \times 10^9$	
	2.50×10^5	$(1.30 \pm 0.18) \times 10^9$	
	5.00×10^5	$(1.16 \pm 0.15) \times 10^9$	
	1.00×10^6	$(1.90 \pm 0.39) \times 10^9$	
HIV-1 _{MN}	1.56×10^4	$(1.46 \pm 0.10) \times 10^9$	$(1.7 \pm 0.13) \times 10^9$
	3.13×10^4	$(1.82 \pm 0.14) \times 10^9$	
	6.25×10^4	$(1.84 \pm 0.14) \times 10^9$	
	1.25×10^5	$(1.65 \pm 0.09) \times 10^9$	
	2.50×10^5	$(1.89 \pm 0.14) \times 10^9$	
	5.00×10^5	$(1.45 \pm 0.15) \times 10^9$	
HIV-2 _{NIH2}	3.13×10^4	$(9.46 \pm 0.64) \times 10^7$	$(8.5 \pm 0.54) \times 10^7$
	6.25×10^4	$(9.16 \pm 0.42) \times 10^7$	
	1.25×10^5	$(8.89 \pm 0.76) \times 10^7$	
	2.50×10^5	$(7.67 \pm 0.38) \times 10^7$	
	5.00×10^5	$(7.48 \pm 0.42) \times 10^7$	
	1.00×10^6	$(8.56 \pm 0.64) \times 10^7$	

For each viral strain and target cell concentration, target cells were infected at seven sCD4 concentrations. For HIV-1_{HXB3}, sCD4 concentrations were: 0, 4.0×10^{-10} , 8.0×10^{-10} , 1.2×10^{-9} , 1.6×10^{-9} , 2.0×10^{-9} , $2.4 \times 10^{-9} \text{ M}$. For HIV-1_{MN}, sCD4 concentrations were 0, 2.0×10^{-10} , 4.0×10^{-10} , 6.0×10^{-10} , 8.0×10^{-10} , 1.0×10^{-9} , $1.2 \times 10^{-9} \text{ M}$. For HIV-2_{NIH2}, sCD4 concentrations were 0, 4.0×10^{-9} , 8.0×10^{-9} , 1.2×10^{-8} , 1.6×10^{-8} , 2.0×10^{-8} , $2.4 \times 10^{-8} \text{ M}$. Using the same soluble CD4 (CD4T) as our study and Scatchard analysis, methods based on the chemical K_{assoc} for HIV-1_{IIIB} was reported as $(1.19 \pm 0.14) \times 10^9 \text{ M}^{-1}$ (ref. 7). Using enzyme-linked immunosorbent assay (ELISA), the chemical K_{assoc} for HIV-1_{IIIB} and HIV-2_{ROD} was $(8.0 \pm 4.8) \times 10^8$ and $(2.2 \pm 1.1) \times 10^7 \text{ M}^{-1}$, respectively⁴¹. HIV-1_{IIIB} and HIV-1_{HXB3} differ by 0.5% in their gp120 nucleotide sequence; HIV-2_{ROD} and HIV-2_{NIH2} differ by 12% (ref. 26). On the basis of nucleotide similarity, the HIV-1 strains permit the most direct comparison between biological and chemical K_{assoc} .

K_{assoc} , giving a numerical value of one to the sCD4 concentration at which half the gp120 molecules are blocked (see Fig. 3). Surprisingly, after normalization, the shape of the inverse infection plot is independent of the target cell concentration, the value of K_{assoc} , and the viral type.

When sCD4 blocks less than half of the gp120 molecules on virus optimized for infectivity²³ ($\beta < 1$, so $1/(1 + \beta) > 1/2$, where $\beta = [\text{sCD4}] \times K_{\text{assoc}}$), the normalized infection plot (Fig. 3) is linear, indicating each gp120 molecule independently and equivalently contributes to infection^{21,22}. But when more than half the gp120 molecules are blocked ($\beta > 1$), the assumption of equivalent and independent gp120 monomers cannot explain all the blocking of infection *in vitro*.

Some explanations can be excluded immediately. Because the cloned viral stock (HIV-1_{HXB3}) gave the same results as the two uncloned viral stocks (HIV-1_{MN} and -2_{NIH2})²⁶, the cooperativity is probably not caused by heterogeneous gp120s with differing K_{assoc} . Assay artefacts causing spurious cooperativity, such as interference by soluble gp120, have been ruled out (see legend to Fig. 1). Soluble CD4 might have enhanced the shedding or direct inactivation of viral gp120, but a simple extension of our model^{21,22} shows the slope of the inverse infection plot would simply overestimate K_{assoc} and would not produce the upward curvature shown in Fig. 3.

One possibility is that a single gp120-CD4 interaction initiates viral binding, whereas viral adsorption and penetration require subsequent recruitment of additional gp120-CD4 interactions. Recruitment accounts for cooperativity and is consistent with experiments using sCD4 and immunoglobulins to inhibit infection after initial binding^{12,17,27}. The recruitment hypothesis predicts that aged viral stocks (with virions that have shed most of

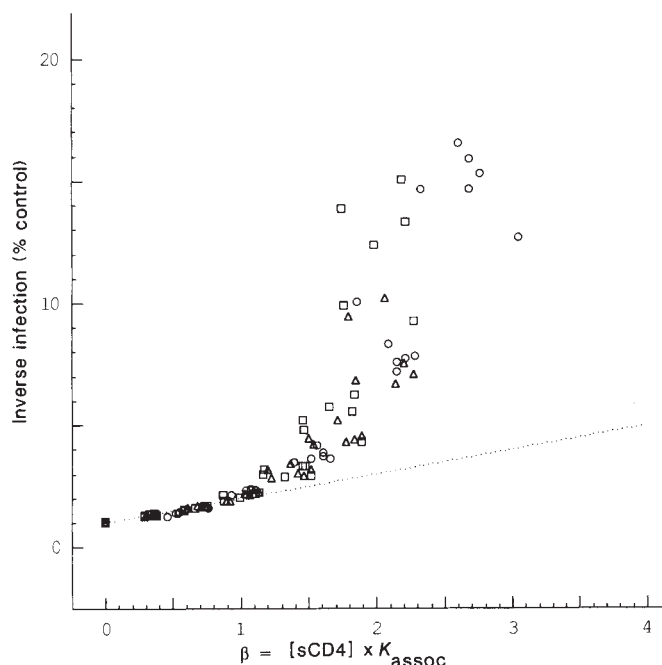


FIG. 3 Analysing the cooperative blocking activity of sCD4. Dividing inverse infection (Fig. 2) by its control (inverse infection without blocker) and multiplying sCD4 concentrations by the apparent K_{assoc} yields a normalized plot of inverse infection. Each normalization used the K_{assoc} calculated from the same target cell concentration (Table 1). The fraction of gp120 molecules that are either free or blocked by a particular concentration of sCD4 is given by $1/(1 + \beta)$ and $\beta/(1 + \beta)$ respectively, where $\beta = [\text{sCD4}] \times K_{\text{assoc}}$. Results at all six target cell concentrations for HIV-1_{IXB3} (○), HIV-1_{MN} (□) and HIV-2_{NIH2} (△). The dotted line represents blocking based on independent and equivalent gp120s (refs 21, 22). When less than half the gp120 molecules are blocked, the points lie on the dotted line. When half the gp120 molecules are blocked, deviations indicating sCD4 blocking synergy begin to occur. The synergies are identical for each viral strain.

their gp120) will be more easily blocked, thus enhancing the cooperativity in Fig. 3. Another explanation is that there are allosteric interactions between gp120s on the viral coat. This, in contrast to the recruitment model, predicts that the cooperativity of aged viral stocks will be unchanged or even decreased.

The upward curvature of the inverse infection plot is apparent only after half of the gp120 binding sites are blocked. Thus, the HIV envelope is covered by a highly redundant number of gp120 molecules which act independently at low sCD4 concentrations. Unlike HIV, viruses such as polio and influenza are covered by interacting (metastable) capsid polypeptide subunits and interacting glycoprotein subunits, respectively, which present relatively few critical neutralization sites. When a fraction of these sites are blocked by neutralizing antibodies, a non-local transition in subunit orientation is induced that inactivates the virus²⁸⁻³⁰. This property may contribute to the humoral efficacy of vaccines against polio and influenza. Neutralizing HIV, however, seems to be fundamentally different. □

Received 19 December 1989; accepted 17 May 1990.

1. Dalgleish, A. G. *et al. Nature* **312**, 763-767 (1984).
2. Klatzmann, D. *et al. Nature* **312**, 767-768 (1984).
3. McDougall, J. S. *et al. Science* **231**, 382-385 (1986).
4. Kowalski, M. *et al. Science* **237**, 1351-1355 (1987).
5. Bedinger, P. *et al. Nature* **334**, 162-165 (1988).
6. Lasky, L. A. *et al. Cell* **50**, 975-985 (1987).
7. Smith, D. H. *et al. Science* **238**, 1704-1707 (1987).
8. Fisher, R. A. *et al. Nature* **331**, 76-78 (1988).
9. Hussey, R. E. *et al. Nature* **331**, 78-81 (1988).
10. Deen, K. C. *et al. Nature* **331**, 82-84 (1988).
11. Trauneker, A., Lücke, W. & Karjalainen, K. *Nature* **331**, 84-86 (1988).
12. Clapham, P. R. *et al. Nature* **337**, 368-370 (1989).

13. Capon, D. J. *et al. Nature* **337**, 525-531 (1989).
14. Trauneker, A., Schneider, J., Kiefer, H. & Karjalainen, K. *Nature* **339**, 68-70 (1989).
15. Nara, P. L., Hwang, K. M., Rausch, D. M., Lifson, J. D. & Eiden, L. E. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7139-7143 (1989).
16. Sun, N.-C. *et al. J. Virology* **63**, 3579-3585 (1989).
17. Byrn, R. A. *et al. J. Virology* **63**, 4370-4375 (1989).
18. Watanabe, M. *et al. Nature* **337**, 267-270 (1989).
19. Schooley, R. T. *et al. Ann. intern. Med.* **112**, 247-253 (1990).
20. Kahn, J. O. *et al. Ann. intern. Med.* **112**, 254-261 (1990).
21. Layne, S. P., Spouge, J. L. & Dembo, M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4644-4648 (1989).
22. Spouge, J. L., Layne, S. P. & Dembo, M. *Bull. math. Biol.* **51**, 715-730 (1989).
23. Nara, P. L. *et al. AIDS Res. hum. Retroviruses* **3**, 283-302 (1987).
24. Nara, P. L. & Fischinger, P. J. *Nature* **332**, 469-470 (1988).
25. Sweet, R. *et al. Fifth Int. Conf. AIDS Abstr. W.C.O.* **12** (1989).
26. Myers, G. *et al. Human Retroviruses and AIDS Section III*, 9-12 (Los Alamos National Laboratory, New Mexico, 1990).
27. Nara, P. L. in *Vaccines 89* (eds Lerner, R. A., Ginsberg, H., Chanock, R. M. & Brown, F.) 137-144 (Cold Spring Harbor Laboratory, New York, 1989).
28. Icenogle, J. *et al. Virology* **127**, 412-425 (1983).
29. Emini, E. A., Ostapchuk, P. & Wimmer, E. *J. Virology* **48**, 547-550 (1983).
30. Taylor, H. P., Armstrong, S. J. & Dimmock, N. J. *Virology* **159**, 288-298 (1987).
31. Goudsmit, J. *et al. Proc. UCLA Symposia* (in the press).
32. Looney, D. J. *et al. Science* **241**, 357-359 (1988).
33. Saag, M. S. *et al. Nature* **334**, 440-444 (1989).
34. Cordonnier, A., Montagnier, L. & Emerman, M. *Nature* **340**, 571-574 (1989).
35. Zwart, G. *et al. Lancet* **i**, 474 (1990).
36. Özel, M., Pauli, G. & Gelderblom, H. R. *Arch. Virology* **100**, 255-266 (1988).
37. Earl, P. L., Doms, R. W. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 648-652 (1990).
38. Pyle, S. W. *et al. AIDS Res. Hum. Retroviruses* **3**, 387-400 (1987).
39. Kendall, M. & Stuart, A. *The Advanced Theory of Statistics* Vol. 2 (Griffin, London, 1979).
40. Efron, B. & Tibshirani, R. *Stat. Sci.* **1**, 54-77 (1986).
41. Moore, J. P. *AIDS* **4**, 297-305 (1990).

ACKNOWLEDGEMENTS. We thank Drs R. C. Gallo, F. Wong-Staal, M. Popovic and D. Zagury for providing HIV-1_{IXB3}, HIV-1_{MN} and HIV-2_{NIH2}, and D. J. Capon and R. Ward (Genentech) for sCD4. This work was financially supported by the US Army Medical Research and Development Command, US Department of Energy, and the NIH.

Additional deletion in sex-determining region of human Y chromosome resolves paradox of X,t(Y;22) female

David C. Page*, Elizabeth M. C. Fisher*, Barbara McGillivray† & Laura G. Brown*

* Whitehead Institute for Biomedical Research, Nine Cambridge Center, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

† University of British Columbia Clinical Genetics Unit, Vancouver, British Columbia V6H 3N1, Canada

WHETHER a human embryo develops as a male or a female is determined by the presence of the Y chromosome^{1,2}. The sex-determining function lies entirely in interval 1A, inasmuch as most XX individuals with descended testes and normal male external genitalia carry this small region of the Y chromosome³. We have localized an essential part of the sex-determining function to a portion of interval 1A, on the basis of the discovery of a female with a reciprocal Y;22 translocation and part of 1A deleted at the translocation breakpoint³. Recently, a paradox has arisen with the report⁴ of four partially masculinized XX individuals who carry only a portion of interval 1A—a portion that does not overlap the deletion in the X,t(Y;22) female. These recent findings imply that the sex-determining function lies in the portion of 1A present in the four XX intersexes and not in the portion deleted in the X,t(Y;22) female. To explain the X,t(Y;22) individual, it was proposed that she was female because of a chromosomal position effect⁴ or delayed development of the gonadal soma⁵. Here we report that the X,t(Y;22) female has a deletion of a second portion of interval 1A—a portion corresponding closely to that present⁴ in the XX intersexes. This resolves the apparent contradiction. Nonetheless, phenotype-genotype correlations suggest that two or more genetic elements in interval 1A may contribute to the sex-determining function of the Y chromosome. The X,t(Y;22) female

lacks the *ZFY* gene but does not exhibit the complex phenotype known as Turner's syndrome, arguing against the hypothesis⁵ that *ZFY* is the Turner's syndrome gene on the Y chromosome.

We have previously cloned, by chromosome-walking, a 230-kilobase (kb) segment of the sex-determining region of the Y chromosome, including part of deletion interval 1A1, all of intervals 1A2 and 1B, and part of 1C (ref. 3). (Intervals 1A1 and 1A2 together constitute interval 1A; see Fig. 1). By continuing this walk, we cloned the remaining, distal portion of interval 1A1; the pseudoautosomal boundary⁶ was crossed.

Various DNA sequences derived from the distal portion of 1A1 were then hybridized to Southern blots of genomic DNAs from individuals known or suspected to carry part but not all of the Y chromosome, including several XX males and XY females. In general, the results were concordant with results previously obtained with probes from proximal 1A1 and 1A2. With one exception, all individuals were either positive for all of interval 1A (both distal and proximal 1A1, as well as 1A2) or negative for all of interval 1A.

The exception was WHT1013, a female with a reciprocal Y;22 translocation. This X,t(Y;22) female was previously shown³ to have a deletion of intervals 1A2 and 1B but otherwise to carry most of the Y chromosome, including proximal 1A1. To our surprise, we found that this X,t(Y;22) female lacks the most distal 50 kb of interval 1A1. For example, she carries an 8-kb Y-specific *TaqI* fragment detected by probe pDP1226 (about 65 kb proximal to the pseudoautosomal boundary) but lacks a 5.5-kb Y-specific *HindIII* fragment detected by probe pDP1225 (about 10 kb proximal to the boundary). As shown in Fig. 1, this deletion divides interval 1A1 into intervals 1A1A (absent in the X,t(Y;22) female) and 1A1B (present in the X,t(Y;22) female). Thus, X,t(Y;22) female WHT1013 has deletions of two portions of the Y chromosome—intervals 1A2 and 1B, and interval 1A1A. On the normal Y chromosome, these two deleted segments are separated by interval 1A1B which measures 90 kb.

Because interval 1A1A is immediately proximal to the pseudoautosomal region, it seemed possible that the second

TABLE 1 Transmission of pseudoautosomal RFLPs to the X,t(Y;22) female

Locus	Probe	Recombination with sex phenotype (%)	Ref.	Transmission from	
				father	mother
<i>DXYS14</i>	29C1	50	14,15	+	+
<i>DXYS20</i>	pDP230	50	16	+	NI
<i>DXYS28</i>	pDP411a	38	16	+	NI*
<i>DXYS15</i>	113D	32	15,17	+	+
<i>DXYS17</i>	601	14	15	—	+
<i>MIC2</i>	pDP1001	2	18,19	NI	NI
	pDP1002				
<i>DXYS77</i>	pDP320a	0	20	—	+

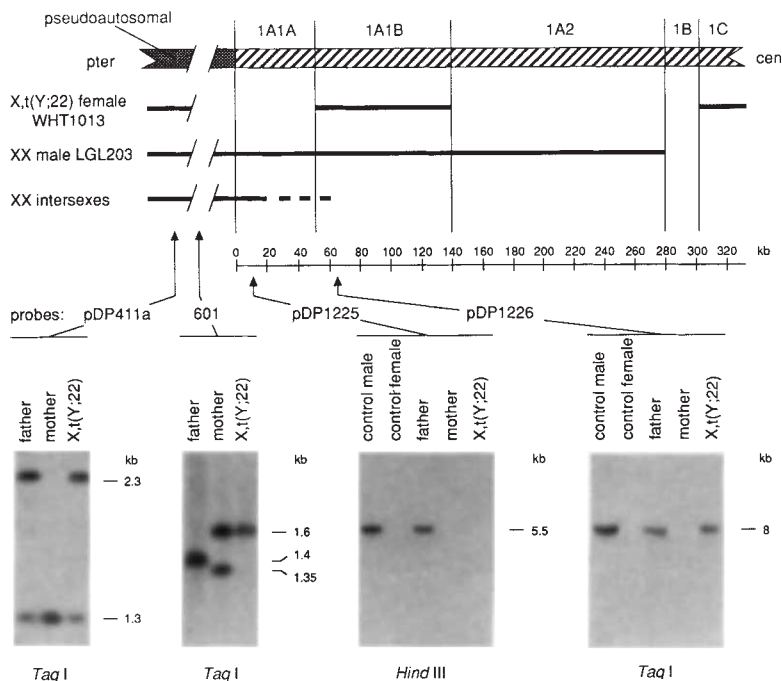
DNA probes detecting pseudoautosomal RFLPs were hybridized to *TaqI*-digested DNAs from X,t(Y;22) female WHT1013 and her parents. The two columns on the right indicate whether or not the X,t(Y;22) female inherited an allele from each of her parents. Loci are ordered from telomere to centromere. (Loci exhibiting more recombination with the sex phenotype during male meiosis are distal to loci exhibiting less recombination.) NI, not informative.

* The *DXYS28* typings (Fig. 1) are consistent with the X,t(Y;22) female having inherited her 1.3-kb allele from her mother.

deletion would extend into the pseudoautosomal region. The single X chromosome in the X,t(Y;22) female is of maternal origin and presumably carries an intact pseudoautosomal region. We wished to determine whether or not the X,t(Y;22) female had inherited a pseudoautosomal region from her father. Accordingly, we traced the transmission of various pseudoautosomal restriction fragment polymorphisms (RFLPs) in the family of the X,t(Y;22) female. We found that she inherited the distal portion of the pseudoautosomal region from her father. At pseudoautosomal locus *DXYS28*, for example, the X,t(Y;22)

FIG. 1 Deletion analysis of X,t(Y;22) female WHT1013 by Southern blotting using pseudoautosomal and Y-specific DNA hybridization probes. Top, schematic representation of the distal short arm of the Y chromosome, oriented with respect to the telomere (pter) and centromere (cen). The pseudoautosomal region undergoes recombination with the X chromosome during male meiosis. Intervals 1A1A, 1A1B, 1A2, 1B and 1C show strictly sex-linked (as opposed to pseudoautosomal) inheritance and are defined by chromosomal deletions depicted immediately below. Black bars indicate the portions of the Y chromosome present in X,t(Y;22) female WHT1013 and in XX male LGL203 (ref. 3; case 6 in ref. 21). The stippled black bar indicates that the breakpoints in four partially masculinized XX individuals ('XX intersexes') fall somewhere between 20 and 60 kb proximal to the pseudoautosomal boundary⁶, in interval 1A1A or 1A1B. Distances from the pseudoautosomal boundary⁶ are shown in kilobases. Bottom, pseudoautosomal probes pDP411a and 601 and Y-specific probes pDP1225 and pDP1226 were hybridized to *TaqI* or *HindIII* digests of DNAs from X,t(Y;22) female WHT1013 and her parents. The sizes in kilobases of the hybridizing restriction fragments are indicated.

METHODS. Plasmids pDP411a and 601 have been described previously^{15,16}. Plasmids pDP1225 and pDP1226 were subcloned from recombinant phages identified by extending a chromosomal walk³ of interval 1A. Plasmid pDP1225 contains a 1.6-kb genomic *HindIII*-*Sall* fragment subcloned into the *HindIII* and *Sall* sites of Bluescript vector. (The *Sall* site in pDP1225 derives from the phage vector and is not present in the human genome.) Plasmid pDP1226 contains a 1.0-kb genomic *HindIII* fragment subcloned into the *HindIII* site of Bluescript vector. Human genomic DNAs were prepared from lymphoblastoid cells, digested with restriction endonuclease, electrophoresed in 0.7% agarose, transferred to nylon mem-



brane, hybridized at 47 °C with ³²P-labelled DNA probes, and exposed with X-ray film, all as previously described³. Before hybridization, radiolabelled inserts of pDP1225 and pDP1226 were preannealed with an excess of sonicated human genomic DNA²².

TABLE 2 Absence of Turner's syndrome features in X,t(Y;22) female

Feature	Turner's syndrome	X,t(Y;22) female
Short stature*	+	—
Lymphoedema	Often	—
Webbing of neck	Often	—
Short fourth metacarpals	Often	—
Increased carrying angle of arms	Often	—
Coarctation of aorta	Often	—
Horseshoe kidney	Often	Normal right kidney Pelvic left kidney

* As compared with parental heights. At age 15 years 10 months, the X,t(Y;22) female's height was 164 cm (60th percentile).

female bears a 2.3-kb allele of paternal origin (Fig. 1, probe pDP411a). (The daughter's 1.3-kb allele is presumably of maternal origin.) But the X,t(Y;22) female did not inherit the proximal portion of the pseudoautosomal region from her father. For example, at *DXYS17*, the X,t(Y;22) female inherited a 1.6-kb allele from her mother but did not inherit the 1.4-kb allele for which her father is homozygous (Fig. 1; probe 601). As summarized in Table 1, the X,t(Y;22) female inherited an allele from her father for loci in the more distal portion of the pseudoautosomal region (*DXYS14*, *DXYS20*, *DXYS28*, and *DXYS15*), but she has no paternal allele for loci in the more proximal portion (*DXYS17* and *DXYS77*).

Thus, the Y chromosomal abnormality in the X,t(Y;22) female is complicated. Compounding the reciprocal Y;22 translocation is the deletion of two portions of the Y chromosome. The first deletion (of intervals 1A2 and 1B) is entirely within Y-specific DNA and is of 160 kb. The second deletion spans 50 kb of Y-specific DNA (interval 1A1A) but extends at least 600 kb and no more than 1,900 kb into the pseudoautosomal region (as estimated from physical maps^{7,8}). Although we suspect that the complex deletion was created during the process of translocation, a firm understanding must await detailed analysis of deletion and translocation breakpoints.

The complex deletion in the X,t(Y;22) female is of immediate relevance to analysis of the sex-determining function of the Y chromosome. Most human XX males carry minute portions of the short arm of the Y chromosome and, though sterile, are otherwise masculinized, with descended testes and normal male external genitalia⁹. Such XX males are heterogeneous with respect to the amount of Y material they carry, but virtually all seem to carry the whole of interval 1A. Indeed, the smallest known segment of the Y chromosome in any such fully masculinized individual is interval 1A, as found in XX male LGL203 (Fig. 1 and ref. 3). Thus, all of the sex-determining function of the Y chromosome, whether one or several genes, maps to 1A (ref. 3). Four partially masculinized XX individuals⁴ carry only a fraction of interval 1A (corresponding roughly to interval 1A1A; Fig. 1), indicating that a substantial portion of the sex-determining function of the Y chromosome maps to interval 1A1A. Having detected a seemingly simple deletion in X,t(Y;22) female WHT1013, we had previously concluded³ that an essential portion of the sex-determining function occurred in interval 1A2. The finding of the second deletion (of interval 1A1A) in the X,t(Y;22) female undermines this previous conclusion but again emphasizes the importance of interval 1A, and especially of 1A1A, in sex determination. Given that four XX individuals⁴ carrying only a small part of 1A are less masculinized than XX individuals carrying all of 1A, it may be that a gene or genes present in the intersexes⁴ and a gene or genes in the remainder of 1A contribute to the sex-determining function of the Y chromosome. In any case, the female phenotype of the X,t(Y;22)

individual is probably simply due to the absence of one or more sex-determining genes found in intervals 1A1A and 1A2.

The X,t(Y;22) female also sheds light on the aetiology of Turner's syndrome, a complex phenotype¹⁰ often associated with a 45,X karyotype¹. Turner's syndrome—or at least much of its somatic component—is apparently the result of monosomy for a gene or genes common to the X and Y chromosomes¹¹. On the Y chromosome, these 'Turner's gene(s)' seem to be located in deletion intervals 1 or 2 (ref. 12; D.C.P. *et al.*, unpublished results). The *ZFY* gene lies in interval 1A2 of the Y chromosome³ and has a closely related homologue, *ZFX*, on the X chromosome¹³. *ZFY* is deleted in the X,t(Y;22) female³. It has been suggested⁵ that *ZFY* and *ZFX* may be Turner's genes. If this were the case, one would expect the X,t(Y;22) female, who has one copy of *ZFX* and lacks *ZFY*, to have some of the somatic features of Turner's syndrome. That she does not (Table 2) argues against *ZFX* and *ZFY* being Turner's genes. □

Received 16 May; accepted 27 June 1990.

1. Ford, C. E. Miller, O. J., Polani, P. E., de Almeida, J. C. & Briggs, J. H. *Lancet* **1**, 711–713 (1959).
2. Jacobs, P. A. & Strong, J. A. *Nature* **183**, 302–303 (1959).
3. Page, D. C. *et al. Cell* **51**, 1091–1104 (1987).
4. Palmer, M. S. *et al. Nature* **342**, 937–939 (1989).
5. Burgoyne, P. S. *Nature* **342**, 860–862 (1989).
6. Ellis, N. A. *et al. Nature* **337**, 81–84 (1989).
7. Brown, W. R. A. *EMBO J.* **7**, 2377–2385 (1988).
8. Petit, C., Levilliers, J. & Weissenbach, J. *EMBO J.* **7**, 2369–2376 (1988).
9. de la Chapelle, A. *Hum. Genet.* **58**, 105–116 (1981).
10. Turner, H. H. *Endocrinology* **23**, 566–574 (1938).
11. Ferguson-Smith, M. A. *J. med. Genet.* **2**, 142–155 (1965).
12. Levilliers, J., Quack, B., Weissenbach, J. & Petit, C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2296–2300 (1989).
13. Schneider-Gadicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R. & Page, D. C. *Cell* **57**, 1247–1258 (1989).
14. Cooke, H. J., Brown, W. R. A. & Rappold, G. A. *Nature* **317**, 687–692 (1985).
15. Rouyer, F. *et al. Nature* **319**, 291–295 (1986).
16. Page, D. C. *et al. Genomics* **1**, 243–256 (1987).
17. Simmler, M.-C. *et al. Nature* **317**, 692–697 (1985).
18. Goodfellow, P. J., Darling, S. M., Thomas, M. S. & Goodfellow, P. N. *Science* **234**, 740–743 (1986).
19. Page, D. C., Brown, L. G. & de la Chapelle, A. *Nature* **328**, 437–440 (1987).
20. Fisher, E. M. C., Alitalo, T., Luoh, S.-W., de la Chapelle, A. & Page, D. C. *Genomics* **7**, 625–628 (1990).
21. de la Chapelle, A. *et al. in Aspects of Human Genetics* (ed. San Roman, C. & McDermott, A.) 125–142 (Karger, Basel, 1984).
22. Litt, M. & White, R. L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6206–6210 (1985).

ACKNOWLEDGEMENTS. We thank Howard Cooke and Jean Weissenbach for DNA probes, and Eric Lander, Anne Rickey, Peggy Beer-Romero, Elizabeth Simpson and Douglas Vollrath for comments on the manuscript. This work was supported by the NIH, the Dupont Center of Molecular Genetics at the Whitehead Institute, the Lucille P. Markey Charitable Trust, and the Searle Scholars Program/Chicago Community Trust. E.M.C.F. was supported by a fellowship from EMBO.

Negative regulation of transforming growth factor- β by the proteoglycan decorin

Yu Yamaguchi, David M. Mann & Erkki Ruoslahti*

Cancer Research Center, La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, California 92037, USA

DECORIN is a small chondroitin–dermatan sulphate proteoglycan consisting of a core protein and a single glycosaminoglycan chain^{1–3}. Eighty per cent of the core protein consists of 10 repeats of a leucine-rich sequence of 24 amino acids^{2,4}. Similar repeats have been found in two other proteoglycans, biglycan⁵ and fibromodulin⁶, and in several other proteins including *Drosophila* morphogenetic proteins^{7–11}. Expression of high levels of decorin in Chinese hamster ovary cells has a dramatic effect on their morphology and growth properties¹. We now report that this effect is due at least in part to the ability of decorin to bind transforming growth factor- β , an autocrine factor^{12,13} that stimulates the growth

* To whom correspondence should be addressed.

of Chinese hamster ovary cells. As transforming growth factor- β induces synthesis of decorin in many cell types^{14,15}, our results suggest that decorin may be a component of a feedback system regulating cell growth.

We found that culture medium from decorin-expressing Chinese hamster ovary (CHO) cells contained growth-inhibitory activity as measured by a [3 H]thymidine incorporation assay. The medium from control cell lines lacked such activity.

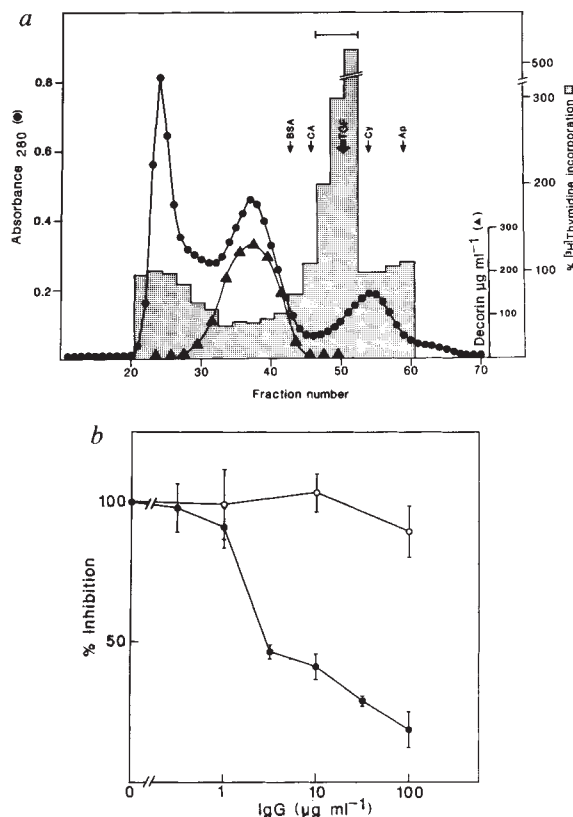


FIG. 1 *a*, Separation of growth-inhibitory activity from decorin-expressing CHO cells by gel filtration on Sepharose CL-6B. Serum-free conditioned medium from decorin overexpressor cells was fractionated by DEAE-Sepharose chromatography in a neutral Tris-HCl buffer and fractions containing growth-inhibitory activity were fractionated further on Sepharose CL-6B under dissociating conditions. The fractions were analysed for protein content, decorin content and growth-regulatory activities. Elution positions of marker proteins are indicated by arrows. BSA, bovine serum albumin ($M_r = 66$ K); CA, carbonic anhydrase (29K); Cy, cytochrome *c* (12.4K); Ap, aprotinin (6.5K); TGF [125 I]TGF- β 1 (25K). *b*, Identification of the growth-stimulatory material from gel filtration as TGF- β . The growth-stimulatory activity from the late fractions from Sepharose CL-6B (bar in *a*) was identified as TGF- β 1 activity by inhibiting the activity with protein A-purified IgG from an anti-TGF- β 1 antiserum¹⁵. As the sequences of various TGF- β s are very similar²⁵, the activity of TGF- β s other than TGF- β 1, if present, could also have been inhibited. Data represent per cent inhibition of growth-stimulatory activity in a [3 H]thymidine incorporation assay. Each point shows the mean $\pm$ standard deviation of triplicate determinations. Anti-TGF- β 1 ($\bullet$); normal rabbit IgG ($\circ$). The 100% inhibition represents 96,000 c.p.m. (0%, 45,000 c.p.m.).

METHODS. A CHO cell line (D61, clone 61 in ref. 1) that overexpresses human decorin as a result of DNA transfection and subsequent gene amplification, and a control transfected cell line (C60, line C in ref. 1) were used. Serum-free conditioned medium prepared as described previously²⁶ was fractionated on DEAE-Sepharose (30 ml of resin per litre of medium), the resin was washed extensively with 50 mM Tris-HCl buffer (pH 7.4) containing 0.2 M NaCl and eluted with a linear gradient of 0.2–1 M NaCl in Tris buffer. Fractions containing growth-inhibitory activity were dialysed, lyophilized, dissolved in 50 mM sodium acetate buffer (pH 5.9) containing 4 M guanidine-HCl and applied to a 1.5×70 -cm column of Sepharose CL-6B equilibrated with the same buffer. A [3 H]thymidine incorporation assay was performed as in Table 1 using an unamplified CHO cell line transfected with pSV2dhfr only (line A in ref. 1, called CHO cells here) as the test cells.

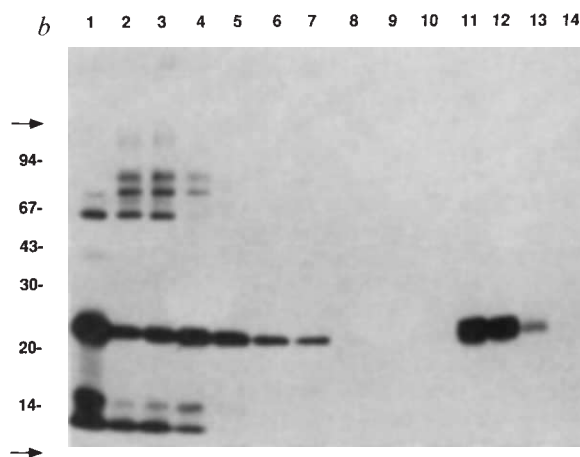
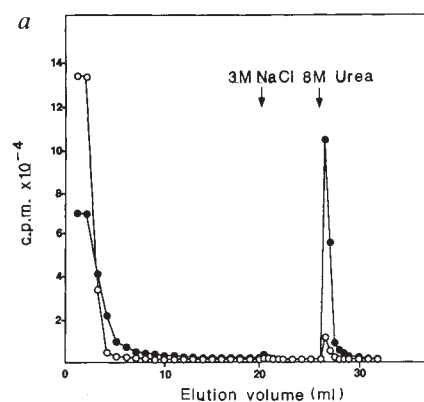


FIG. 2 Binding of [125 I]TGF- β 1 to decorin-Sepharose. *a*, Fractionation of [125 I]TGF- β 1 by decorin-Sepharose affinity chromatography. [125 I]TGF- β 1 (5×10^5 c.p.m.) was incubated in BSA-coated polypropylene tubes with 0.2 ml of packed decorin-Sepharose ($\bullet$) or gelatin-Sepharose ($\circ$) in 2 ml PBS, pH 7.4, containing 1 M NaCl and 0.05% Tween-20. After overnight incubation, the affinity matrices were transferred into BSA-coated disposable columns (Bio Rad) and washed with the binding buffer. Elution was effected, first with 3 M NaCl in the binding buffer and then with 8 M urea in the same buffer. *b*, Analysis of eluents of decorin-Sepharose affinity chromatography by SDS-PAGE under nonreducing conditions. Lane 1, the original [125 I]-labelled TGF- β 1 sample; lanes 2–7, flow-through and wash fractions; lanes 8–10, 3 M NaCl fractions; lanes 11–14, 8 M urea fractions. Arrows, top and bottom of the 12% separating gel.

METHODS. For large-scale purification of recombinant decorin, D61 cells were grown as described previously¹, except that fetal calf serum that had been treated with DEAE-Sepharose (DEAE-FCS) was used in the culture medium. DEAE-FCS was prepared by incubating FCS (500 ml) with DEAE-Sepharose (50 ml) (Pharmacia) that had been pre-washed with PBS buffer and the unbound fraction used in cell culture. Decorin was purified from the culture medium by performing DEAE-Sepharose chromatography twice, essentially as described in Fig. 1, except that the column was washed with buffer containing 6 M urea and 6 mM CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulphonate) before elution. Fractions from the second column—in which decorin was the only material detected in Coomassie blue or toluidine blue-stained SDS-PAGE gels—were pooled and coupled to CNBr-activated Sepharose 4B (Pharmacia) according to the manufacturers instructions. Purified TGF- β 1 from human platelets (Calbiochem, La Jolla, California) was iodinated by a chloramine-T method²⁷. To determine the affinity of the binding of TGF- β 1 to the decorin-Sepharose, 20 μ l of the resin was incubated with increasing concentrations of [125 I]TGF- β 1 in 100 μ l 20 mM Tris buffer, 0.14 M NaCl, 0.1% BSA (pH 7.4) for 4 h at 4 $^{\circ}$ C with continuous mixing. After incubation, the resin was washed four times with cold binding buffer, transferred into fresh tubes and bound-radioactivity was determined. Nonspecific binding was determined in the presence of a 100-fold molar excess of unlabelled TGF- β 1. Specific binding was calculated by subtracting nonspecific binding from total binding and analysed by LIGAND program²⁸.

TABLE 1 Decorin-Sepharose affinity chromatography of unlabelled TGF- β 1 monitored by growth inhibition assay in Mv1Lu cells

Elution	TGF- β 1 (ng)	
	Decorin-Sepharose	BSA-Sepharose
Flow-through and wash	2.7 (2.3%)	82.0 (93.9%)
3 M NaCl	2.2 (1.8%)	1.3 (1.5%)
8 M Urea	116.0 (95.9%)	4.0 (4.6%)

TGF- β 1 (180 ng) was incubated with 0.2 ml packed volume of decorin-Sepharose or BSA-agarose (Sigma) in PBS (pH 7.4) containing 1% BSA. After overnight incubation at 4 °C, the resins were washed with 15 ml of this buffer and eluted, first with 5 ml or 3 M NaCl in PBS buffer then with 5 ml of PBS buffer containing 8 M urea. Aliquots of each pool were dialysed against α -modified minimal essential medium (GIBCO) and assayed for the inhibition of [3 H]thymidine incorporation in Mv1Lu cells²⁴. In brief, dialysed samples were added to a monolayer of Mv1Lu cells in 96-well plates in the presence of 1% FCS and incubated for 24 h. Cells were pulsed for 2 h with [3 H]thymidine (1 μ Ci per well), extracted with 10% trichloroacetic acid and the radioactivity incorporated into acid-insoluble materials was counted. The amount of TGF- β 1 in each pool was calculated from the standard curve of [3 H]thymidine incorporation obtained from a parallel experiment with known concentrations of TGF- β 1. The partial recovery of the TGF- β 1 activity may be due to loss of TGF- β 1 in the dialyses.

Fractionation of the inhibitory medium by ion-exchange chromatography on DEAE-Sepharose yielded a single peak of growth-inhibitory activity that eluted with 0.35–0.55 M NaCl and that coincided with decorin. Further fractionation of the DEAE-bound growth-inhibitory activity by gel filtration in the presence of guanidine-HCl demonstrated that the inhibitory activity copurifies with decorin (Fig. 1a). Unexpectedly, this fractionation also revealed a peak of growth-stimulatory activity at relative molecular mass about 25,000 ($M_r \sim 25K$). These results indicated that the growth-stimulatory activity had bound

to the DEAE column through decorin and dissociated from it in the presence of guanidine-HCl in the subsequent fractionation. Because the size of the decorin-associated growth-stimulatory activity was similar to that of transforming growth factor- β 1 (TGF- β 1), we suspected that the activity might be due to this growth factor.

When the effect of the decorin-associated 25K fraction was compared with that of purified TGF- β 1, the same response was obtained in each of the three cell lines studied. Both substances were growth-stimulatory in CHO cells and mouse 3T3-L1 cells, but inhibitory in mink Mv1Lu cells (data not shown). Furthermore, an anti-TGF- β 1 antibody¹⁵ neutralized these activities in CHO cells (Fig. 1b). These results indicate that the growth-regulatory effect of the 25K fraction is due to the presence of TGF- β , which was active only after it was released from binding to decorin.

Affinity chromatography on decorin-Sepharose showed that 25–30% of applied [125 I]-labelled TGF- β 1 was bound and could be eluted with 8 M urea, but not with 3 M NaCl (Fig. 2a). SDS-PAGE analysis of the fractions showed that among the several bands seen in the original [125 I]-labelled TGF- β 1 preparation, only the main band at 25K that represents the intact TGF- β 1 dimer bound to the decorin-Sepharose (Fig. 2b). The portions of the 25K material that did not bind, or that were only retarded by the column, may represent TGF- β 1 denatured by the iodination, as essentially all TGF- β activity from an unlabelled TGF- β 1 preparation bound to the decorin-Sepharose in a similar fractionation (Table 1). A dissociation constant (K_d) of 1.5×10^{-9} M was derived from these binding studies for the TGF- β 1-decorin interaction.

In a microtitre assay, recombinant decorin and decorin isolated from bovine skin inhibited the binding of [125 I]-labelled TGF- β 1 to decorin-coated wells (Fig. 3a). It is of interest that biglycan, which is closely related to decorin⁶, was as effective an inhibitor as decorin. Chicken cartilage proteoglycan, which

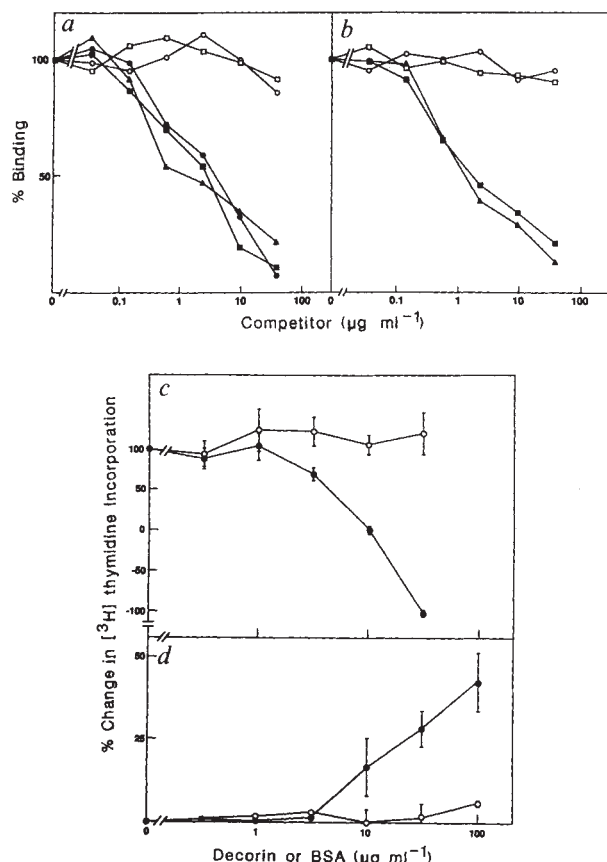


FIG. 3 *a*, Competition of [125 I]TGF- β 1 binding to decorin-coated microtitre wells by recombinant decorin (●), decorin isolated from bovine skin²¹ (PGII) (■), biglycan isolated from bovine articular cartilage²¹ (PGI) (▲), chicken cartilage proteoglycan²⁹ (○), and BSA (□). Each point represents the mean of duplicate samples. *b*, Competition of [125 I]TGF- β 1 binding with chondroitinase ABC-treated proteoglycans and BSA. Concentrations of competitors were expressed as intact proteoglycan. Symbols are as in *a*. *c*, Inhibition of TGF- β 1-induced proliferation of CHO cells by decorin. [3 H]thymidine incorporation assay was performed as described, in the presence of 5 ng ml⁻¹ of TGF- β 1 and the indicated concentrations of purified decorin (●) or BSA (○). The data represent per cent neutralization of growth stimulation; [3 H]thymidine incorporation in the absence of either TGF- β 1 or decorin is 0% (40,000 c.p.m.); incorporation in the presence of TGF- β 1 but not decorin is 100% (65,000 c.p.m.). Each point shows the mean $\pm$ s.d. of triplicate samples. *d*, Neutralization of TGF- β 1-induced growth inhibition of Mv1Lu cells by decorin. Assay was performed with Mv1Lu cells as described in *c*, except that TGF- β 1 was added at 0.5 ng ml⁻¹. The data represent per cent neutralization of TGF- β -induced growth inhibition; [3 H]thymidine incorporation in the presence of neither TGF- β 1 or decorin is 100% (126,000 c.p.m.); incorporation in the presence of TGF- β 1 but not decorin is 0% (68,000 c.p.m.). Symbols as in *c*.

METHODS. *a*, *b*, The wells of a 96-well microtitre plate were coated overnight with 2 μ g ml⁻¹ of recombinant decorin in 0.1 M sodium carbonate buffer, pH 9.5. The samples containing 5×10^4 c.p.m. of [125 I]TGF- β 1 and various concentrations of competitors in PBS containing 0.05% Tween-20 were added to each well, and the plates incubated at 37 °C for 4 h (at 4 °C overnight in experiments with chondroitinase ABC-digested proteoglycans). The bound-radioactivity was solubilized with 1% sodium dodecyl sulphate in 0.3 M NaOH. Total binding without competitors was $\sim$ 4% under the conditions used. Nonspecific binding, determined by adding 100-fold molar excess of unlabelled TGF- β 1 over the labelled TGF- β 1 to the incubation mixture, was about 13% of total binding. For the preparation of core proteins, proteoglycans were digested with chondroitinase ABC (Seikagaku, Tokyo) as described³⁰.

carries many chondroitin sulphate chains but has an entirely unrelated core protein¹⁶, did not cause any inhibition. Furthermore, recombinant decorin core protein that lacks glycosaminoglycans¹⁷ (not shown) and chondroitinase-digested decorin and biglycan were inhibitory, whereas chondroitinase-digested cartilage proteoglycan was not (Fig. 3b). These results suggest that TGF- β 1 binds to the core protein of decorin and biglycan.

The ability of decorin to modulate the activity of TGF- β 1 was examined in [³H]thymidine-incorporation assays. In CHO cells decorin neutralized the growth-stimulatory activity of TGF- β 1 and suppressed it below the level observed without any added TGF- β 1 (Fig. 3c). This additional effect on CHO cells may be due to the inhibition of endogenous TGF- β , as we found 0.25 ng ml⁻¹ of TGF- β activity in serum-free medium of these cells after 24 h in culture (as determined by growth inhibition on the Mv1Lu cells, result not shown). Thus, it seems that TGF- β serves as an autocrine growth factor in these particular cells, and that the effect of added decorin—and that of decorin complementary DNA expression¹—occurs through suppression of this autocrine activity. In the Mv1Lu cells, which respond to TGF- β by growth inhibition, the addition of decorin partially reversed this inhibitory effect (Fig. 3d). These opposing effects of decorin indicate that the mechanism of decorin action is the neutralization of TGF- β activity through binding of the growth factor.

The results reported here show that decorin binds TGF- β and thereby modulates the activity of this growth factor. An increasing number of observations indicate that proteoglycans can regulate cell proliferation through interaction of various growth factors (see ref. 18 for review). One of the distinguishing features of the decorin-TGF- β interaction is that the binding occurs through the core protein, not the glycosaminoglycan chain. The leucine-rich repeats that characterize decorin and biglycan may be the site of TGF- β binding in these proteoglycans.

TGF- β stimulates the synthesis of decorin and biglycan^{14,15}. Thus, the ability of decorin to neutralize TGF- β activity suggests that these proteoglycans serve as negative-feedback regulators of TGF- β activity. One of the TGF- β receptors is a cell-surface proteoglycan that also exists in a soluble form¹⁹. This proteoglycan, called betaglycan, seems to be different from the ones we have studied here because its core protein is much larger than that of decorin and biglycan¹⁹. It is not known whether binding of TGF- β to betaglycan inhibits the activity of the growth factor.

As decorin and biglycan are extracellular matrix components of various connective tissues *in vivo*^{20,21}, the interaction of TGF- β with these proteoglycans may bind the growth factor to the extracellular matrix. Another possible fate of TGF- β -decorin complexes is endocytosis through a receptor that recognizes decorin²². Finally, neutralization of TGF- β activity with decorin might allow therapeutic intervention in cases where excessive TGF- β activity causes a fibrotic condition¹³ or when TGF- β serves as an autocrine growth factor for tumour cells²³. □

18. Ruoslahti, E. *J. biol. Chem.* **264**, 13369–13372 (1989).
19. Andres, J. L., Stanley, K., Chieffetz, S. & Massagué, J. *J. Cell Biol.* **109**, 3137–3145 (1989).
20. Voss, B., Glössl, J., Cully, Z. & Kresse, H. *J. Histochem. Cytochem.* **34**, 1013–1019 (1986).
21. Rosenberg, L. C. *et al. J. biol. Chem.* **260**, 6304–6313 (1985).
22. Hoppe, W., Ranch, U. & Kresse, H. *J. biol. Chem.* **263**, 5926–5932 (1988).
23. Nugent, M. A., Lane, E. A., Keski-Oja, J., Moses, H. L. & Newman, M. J. *Cancer Res.* **49**, 3884–3890 (1989).
24. Chieffetz, S. *et al. Cell* **48**, 409–415 (1987).
25. Derynck, R. *et al. EMBO J.* **7**, 3737–3743 (1988).
26. Fisher, R. A. *et al. Nature* **331**, 76–78 (1988).
27. Frolik, C. A., Wakefield, L. M., Smith, D. M. & Sporn, M. B. *J. biol. Chem.* **259**, 10995–11000 (1984).
28. Munson, P. J. & Rodbard, D. *Analyt. Biochem.* **107**, 220–239 (1980).
29. Goetinck, P. F. in *The Glycoconjugates* Vol. 3 (ed. Horowitz, M. I.) (Academic, New York, 1982).
30. Oike, Y., Kimata, K., Shinomura, T., Nakazawa, K. & Suzuki, S. *Biochem. J.* **191**, 193–207 (1980).

ACKNOWLEDGEMENTS. We thank Dr D. Altieri for affinity calculations, Dr E. Engvall for comments on the manuscript and T. Wrubleski for its preparation. PGI and PGI²¹ were generous gifts of Dr L. Rosenberg (Montefiore Medical Center, New York). Chicken cartilage proteoglycan²⁹ was provided by Dr P. Goetinck at this Institute. This work was supported by the National Cancer Institute, DHHS and postdoctoral fellowships from the Cancer Research Institute (Y.Y.) and the National Cancer Institute (D.M.M.).

Antisense gene that inhibits synthesis of the hormone ethylene in transgenic plants

A. J. Hamilton, G. W. Lycett & D. Grierson

AFRC Research Group in Plant Gene Regulation,
Department of Physiology and Environmental Science,
University of Nottingham School of Agriculture, Sutton Bonington,
Loughborough LE12 5RD, UK

ETHYLENE controls many physiological and developmental processes in higher plants, including ripening of fruit, abscission, senescence and responses to wounding¹. Although the accumulation of messenger RNAs in ripening fruit and senescing leaves has been correlated with ethylene production and perception^{2–4}, the regulatory mechanisms governing ethylene synthesis and the stimulation of gene expression by ethylene are not understood. We have previously shown that the complementary DNA, pTOM13, corresponds to an mRNA whose synthesis is correlated with that of ethylene in ripening fruit and wounded leaves^{5,6,8}. The pTOM13 mRNA encodes a protein of relative molecular mass 35,000⁶. The cDNA and three related genomic clones have been sequenced, but the function of the protein is unknown^{7–9}. We show here that antisense RNA, which has previously been used only to reduce the expression of genes of known function^{10–12}, when applied to pTOM13, reduces ethylene synthesis in a gene dosage-dependent manner. Analysis of these novel mutants suggests that pTOM13 encodes a polypeptide involved in the conversion of 1-aminocyclopropane-1-carboxylic acid to ethylene by the ethylene-forming enzyme (ACC-oxidase).

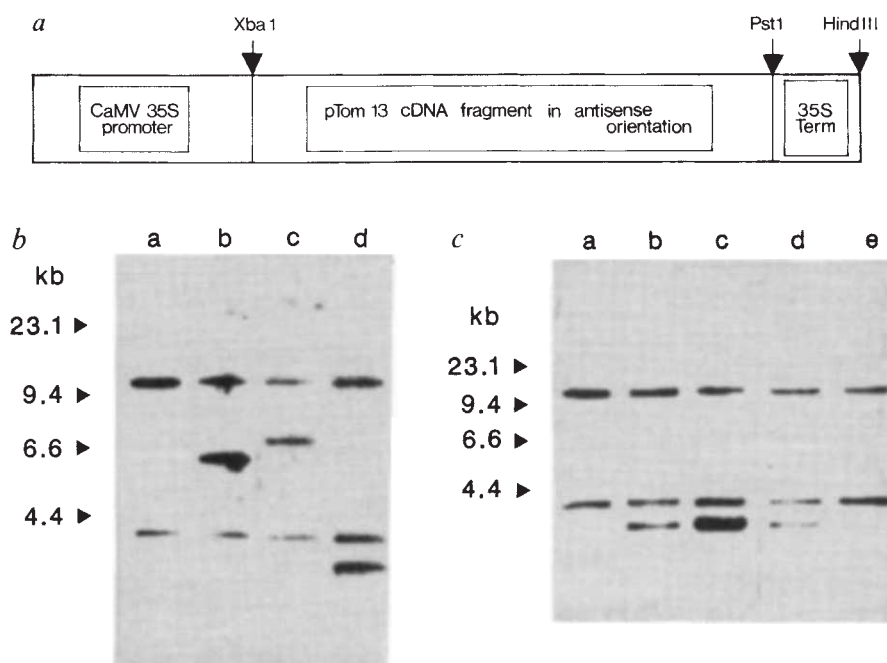
A chimaeric pTOM13 antisense gene, consisting of a 1.1-kilobase (kb) fragment of the cDNA inserted in the reverse orientation between the CaMV 35S promoter and terminator of pDH51¹³, was ligated into BIN19¹⁴ (Fig. 1) and transferred by triparental mating to *Agrobacterium tumefaciens*, which was used to transform tomato plants¹⁵ (*Lycopersicon esculentum* var. Ailsa Craig). Transgenic plants were selected on the basis of kanamycin resistance^{11,13} and mature plants regenerated and grown in compost. The antisense pTOM13 gene was detected by Southern-blot analysis of transformants (Fig. 1). All transformants analysed showed reduced ethylene synthesis. The greatest reduction was in the plant designated C₁, which was chosen for further analysis. Antisense transcripts were measured by probing northern blots of leaf RNA with probes specific for the antisense strand (Fig. 2). Two transcripts were detected (Fig. 2b); the smaller one corresponds to the predicted size of the full-length transcript (1.3 kb) and the larger one is assumed to arise from transcription through the CaMV 35S transcription terminator. In transgenic plants expressing pTOM13 antisense RNA, the accumulation of pTOM13 mRNA in wounded leaves was greatly

Received 29 December 1989; accepted 30 April 1990.

1. Yamaguchi, Y. & Ruoslahti, E. *Nature* **336**, 244–246 (1988).
2. Ruoslahti, E. A. *Rev. Cell Biol.* **4**, 229–255 (1988).
3. Krusius, T. & Ruoslahti, E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7683–7687 (1986).
4. Paththy, L. *J. molec. Biol.* **198**, 567–577 (1987).
5. Fisher, L. W., Termine, J. D. & Young, M. F. *J. biol. Chem.* **264**, 4571–4576 (1989).
6. Oldberg, Å., Antonsson, P., Lindblom, K. & Heinegård, D. *EMBO J.* **8**, 2601–2604 (1989).
7. Takahashi, N., Takahashi, Y. & Putnam, F. W. *Proc. natn. Acad. Sci. U.S.A.* **82**, 73–77 (1985).
8. Titani, F., Takio, K., Handa, M. & Ruggeri, Z. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1906–1910 (1987).
9. Kataoka, T., Broek, D. & Wigler, M. *Cell* **43**, 493–505 (1985).
10. Hashimoto, C., Hudson, K. L. & Young, M. F. *J. Cell Biol.* **52**, 269–279 (1988).
11. Reirke, R., Krantz, D. E., Yen, D. & Zipursky, S. L. *Cell* **52**, 291–301 (1987).
12. Keski-Oja, J., Leof, E. B., Lyons, R. M., Coffey, Jr, J. R. & Moses, H. L. *J. cell. Biochem.* **33**, 95–107 (1987).
13. Roberts, A. B. & Sporn, M. B. in *Handbook exp. Pharmac.* (eds Sporn, M. B. & Roberts, A. B.) Vol. 95, 419–472 (Springer, Heidelberg, 1990).
14. Bassols, A. & Massagué, J. *J. biol. Chem.* **263**, 3039–3045 (1988).
15. Okuda, S., Languino, L. R., Ruoslahti, E. & Border, W. A. *J. clin. Invest.* (in the press).
16. Doege, K., Sasaki, M., Horigan, E., Hassell, J. R. & Yamada, Y. *J. biol. Chem.* **262**, 17757–17767 (1987).
17. Mann, D. M., Yamaguchi, Y., Bourdon, M. A. & Ruoslahti, E. *J. biol. Chem.* **265**, 5317–5323 (1990).

FIG. 1 Structure, integration and inheritance of the pTOM13 antisense gene. **a**, Composition of the antisense gene. The 1.3-kb *Pst*I fragment of the cDNA clone, pTOM13, was ligated into pGEM3 (Promega Biotech). About 0.2 kb of the 3' untranslated region, which is homologous to the 5' end owing to a cloning artefact (A.J.H., unpublished results), was deleted by digestion with *Acc*I, filling in of the overhangs with Klenow enzyme and religation of the vector to create pTRC. A *Pst*I-*Xba*I fragment containing the first 1.1 kb of the cDNA was ligated into the plant expression vector, pDH51¹³, in the antisense orientation. The antisense gene on a *Pvu*II-*Hind*III fragment was inserted into the plant transformation vector pBIN19¹⁴ to create pBASC. The structure of pBASC was confirmed by restriction analysis and Southern hybridization. The vector was transferred from *Escherichia coli* HB101 to *A. tumefaciens* LBA4404 by triparental mating with pRK2013 in *E. coli* HB101¹⁴. Restriction enzyme and hybridization analysis confirmed that the construct was intact. **b**, Genomic Southern blot of *Hind*III-digested DNA from transformed and untransformed tomato plants hybridized to pTOM13 cDNA insert. In both normal (lane a) and transformed plants (lanes b to d), the probe detects both the endogenous 'pTOM13 gene' (at ~10 kb) and an additional closely related gene (at ~4 kb). *Hind*III is expected to cut the antisense gene at the 3' end (see a) and within the plant genome at a position beyond the 5' end of the gene dependent on the site of integration. This results in a new band in transformed plants (lanes b, c and d) but not in the untransformed control plant (lane a). Lane d corresponds to plant line C₁ which was used for further study. **c**, Genomic Southern blot of *Hind*III-digested DNA from selected F₁ progeny of a self-fertilized C₁ plant with 0 (lane a), 1 (lane b) and 2 antisense genes (lane c), the original parent C₁ (lane d) and untransformed control (lane e). As in b, all plants show two bands hybridizing to the two endogenous genes.

METHODS. Genomic DNA was extracted from frozen leaves as described¹⁸.



Approximately 5 µg genomic DNA was digested with *Hind*III, separated on a 0.8% agarose gel and blotted as described¹⁹. Membranes were prehybridized in 5×SSPE²⁰, 0.5% SDS, 5×Denhardt's²⁰ solution and 100 µg ml⁻¹ sheared denatured salmon-sperm DNA at 65 °C for 1 h and hybridized with random-prime-labelled pTOM13 insert in the same buffer for 18 h at 65 °C. The membranes were washed in 0.1×SSPE²⁰, 0.1% SDS at 65 °C and autoradiographed. Molecular size markers are from DNA cut by *Hind*III, electrophoretically separated on the same gel, cut out and stained with ethidium bromide.

reduced compared with controls and was undetectable in ripening fruit (Fig. 2a, c). In these primary transformants the production of ethylene by wounded leaves was inhibited by 68% and in ripening fruit by 87% (Fig. 3). It is of interest that in wounded leaves, where the pTOM13 gene is induced, the dramatic reduction in the accumulation of the sense gene transcript coincided with the almost complete disappearance of the antisense transcript (Fig. 2). An even more pronounced effect was observed in ripening fruit. Not only were sense transcripts not detected in ripening fruit from antisense plants (Fig. 2c), but antisense transcripts were also not detected in fruit from both normal and antisense plants (data not shown). These findings are consistent with results from other antisense systems¹⁰⁻¹², where the levels of mRNA and antisense RNA are either greatly reduced or undetectable. The mechanism of action of this intriguing mutual gene inactivation is unclear but could be explained either by an RNA-DNA interaction interfering with transcription, or a very rapid degradation of a sense RNA/antisense RNA hybrid.

The pTOM13 antisense gene segregated in a population of plants grown from seed derived from a selfed transformant, yielding progeny with 0, 1 or 2 copies of the inserted DNA (Fig. 1b). Fruit of plants that inherited no antisense gene produced similar levels of ethylene to the controls and ethylene evolution in response to wounding was also normal (data not shown). In plants that inherited two antisense genes, ethylene production in ripening fruit was inhibited by 97% (Fig. 3b). In these fruit, colour change was initiated at about the normal time. The extent of reddening was reduced, however, and when stored for several weeks at room temperature, antisense fruit were noticeably more resistant to overripening and shrivelling than controls, although the steady-state levels of RNA and enzyme

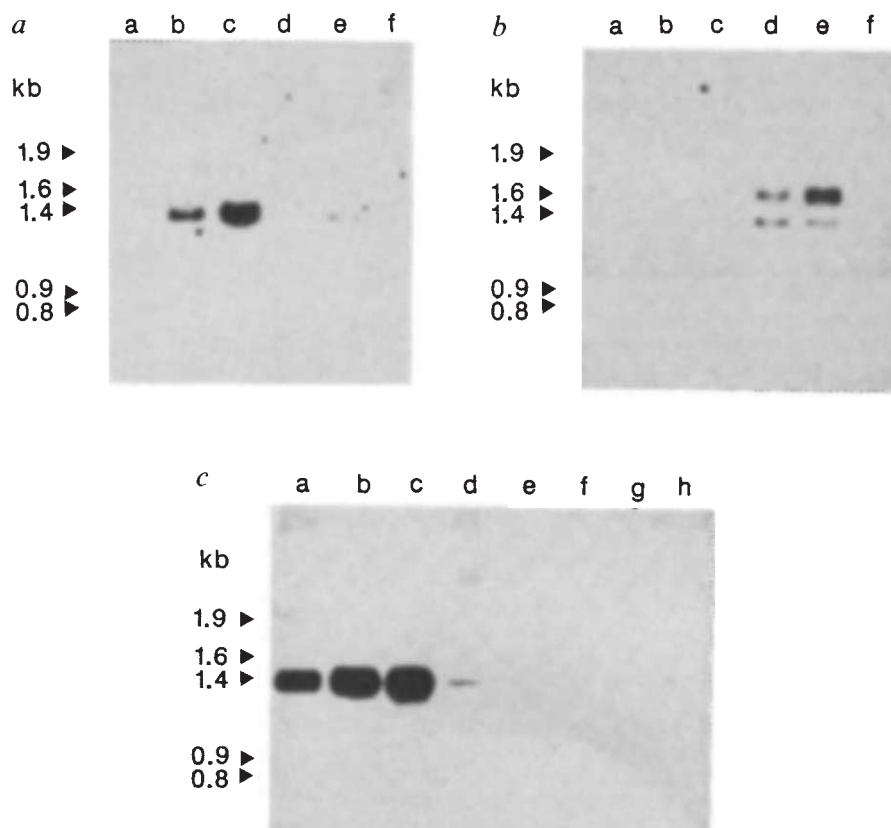
activity of the cell wall degrading enzyme polygalacturonase were similar in fruit from normal and homozygous pTOM13 antisense plants.

These results strongly suggest that the protein encoded by pTOM13 is involved in ethylene synthesis, either by catalysing one of the steps in biosynthesis or by serving a regulatory function, or is involved in recycling the sulphur required for methionine regeneration¹⁶. The two enzymes specific to the pathway are ACC synthase (which catalyses the conversion of S-adenosyl methionine to 1-amino-cyclopropane-1-carboxylic acid, ACC) and ACC oxidase (which converts ACC to ethylene)¹⁶, but despite their importance, no sequence information is available. Evidence has been presented that the *Cucurbita* ACC synthase *in vitro* translation product has a relative molecular mass of 53,000 (*M*_r53K)¹⁷, but pTOM13 encodes a 35K protein⁵, indicating that it does not code for this enzyme. We tested the possibility that ACC oxidase was affected by pTOM13 antisense genes by assaying enzyme activity in wounded leaves. ACC-oxidase activity was inhibited by 66% in hemizygous—and 93% in homozygous—wounded leaves of pTOM13 antisense plants (Fig. 4). Although further evidence will be required to establish the role of the pTOM13 polypeptide unequivocally, the predicted amino-acid sequence of the pTOM13 polypeptide shows 33% identity and 58% similarity to flavone-3-hydroxylase encoded by the *incolorata* gene from *Antirrhinum majus* (A. Prescott and C. Martin, personal communication). This supports the suggestion that it could be part of the ACC-oxidase system as the first step of one proposed reaction mechanism involves the hydroxylation of ACC¹⁶.

These results demonstrate the potential of antisense techniques for assigning functions to gene sequences. The generation

FIG. 2 Expression of pTOM13 sense and antisense RNA. *a*, Wound-induced expression of pTOM13 sense RNA in normal and antisense plants. Samples (10 μ g) of total RNA per lane, extracted from leaves of untransformed (lanes a–c) and C_1 transformed (lanes d–f) plants 0 (lanes a, d), 1 (lanes b, e) and 2 h (lanes c, f) after wounding, were hybridized to a sense-specific pTOM13 RNA probe. *b*, Expression of pTOM13 antisense RNA in normal and antisense plants. RNA (total of 10 μ g per lane), extracted from leaves of untransformed (a–c) and C_1 transformed (d–f) plants 0 (a and d), 1 (b and e) and 2 h (c and f) after wounding, was hybridized to an antisense-specific pTOM13 RNA probe. *c*, Expression of pTOM13 sense RNA during ripening of fruit from normal and antisense plants. RNA (total of 10 μ g per lane), extracted from fruit of untransformed (a–d) and C_1 transformed (e–h) plants at 0 (a and e), 1 (b and f), 3 (c and g) and 5 (d and h) days after the onset of colour change, was hybridized to a sense-specific pTOM13 RNA probe.

METHODS. RNA was extracted from tomato fruit pericarp and leaves as described for leaves⁵ except: (1) contaminating carbohydrate and polyphenolic compounds were removed by differential CTAB (cetyltrimethylammonium bromide) precipitation²¹; (2) DNA was removed by incubation of the final nucleic acid preparation with 100 U of RNase-free DNase (BCL) at 25 °C for 10 min, deproteinized by phenol–chloroform extraction and precipitated twice with 2.5 M ammonium acetate and 2.5 volumes of ethanol. RNA was measured by spectrophotometry and fractionated on 1.2% agarose after denaturing with glyoxal. RNA was blotted onto Hybond N (Amersham) and the membranes prehybridized in $5 \times$ SSPE²⁰, 0.5% SDS, $5 \times$ Denhardt's solution²⁰ and 100 μ g ml⁻¹ sheared denatured salmon sperm DNA. The RNA was hybridized, in the same buffer, to single-stranded probes transcribed from the T7 promoter of pGEM3 (Promega). Membranes were washed in $0.2 \times$ SSPE²⁰,



0.1% SDS at 65 °C and autoradiographed. M_r markers were from DNA cut by *Hind*III and *Eco*RI, boiled and then treated as for RNA samples, electrophoretically separated on the same gel, cut out and stained with ethidium bromide.

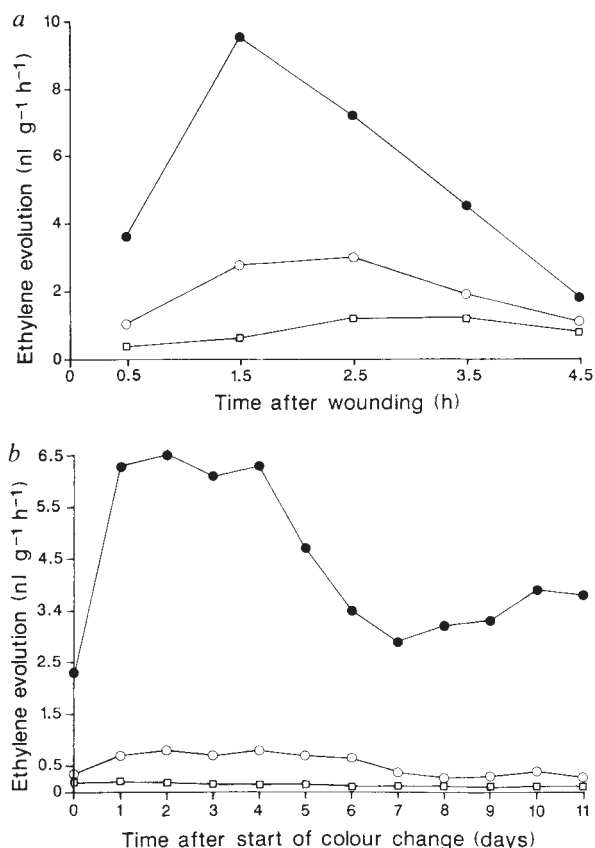


FIG. 3 Ethylene production in normal and antisense plants. *a*, Inhibition of wound ethylene production by pTOM13 antisense genes. Ethylene evolution from untransformed (●), hemizygous (○) and homozygous (□) leaf disks was measured at intervals of 1 h over a 5-h period after wounding. In three paired *t*-tests, the 1–2 h mean values for untransformed with hemizygous; untransformed with homozygous and hemizygous with homozygous were significantly different from each other ($P < 0.001$, $P < 0.001$ and $P < 0.01$, respectively). *b*, Inhibition of ethylene production in ripening fruit by pTOM13 antisense genes. Ethylene evolution from detached fruit of untransformed (●), hemizygous (○) and homozygous (□) plants was measured daily from the onset of colour change for 11 days. In three paired *t*-tests, the 2-day mean values for untransformed with hemizygous, untransformed with homozygous and hemizygous with homozygous were significantly different from each other ($P < 0.05$, $P < 0.01$ and $P < 0.01$ respectively).

METHODS. The tissue or organ to be assayed was placed in a sealed container of known volume, and 1 ml gas sample was taken after 1 h and analysed by gas chromatography as described²².

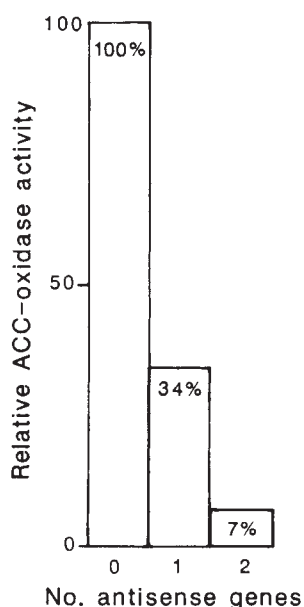


FIG. 4 Levels of ACC-oxidase activity in leaf disks of plants with 0, 1 or 2 pTOM13 antisense genes. ACC-oxidase activities between 1 and 2 h after wounding are shown in the histogram as a percentage of the normal level. In three paired *t*-tests, the mean values for untransformed with hemizygous, untransformed with homozygous and hemizygous with homozygous were significantly different from each other ($P < 0.001$ for each comparison). METHODS. ACC-oxidase activity was assayed by measuring the conversion of exogenous ACC to ethylene in a manner similar to that described²³. Leaf disks were cut with a cork borer and immediately infiltrated under vacuum for 2 min with a solution of 10 mM amino-oxyacetic acid (Sigma), 10 mM 1-amino-cyclopropane-1-carboxylic acid (Sigma) and 100 mM sodium phosphate (pH 6.5). The disks were sealed in 5-ml glass bottles 1 h later and ethylene in the headspace was measured after a further hour by gas chromatography as described²².

of plants in which ethylene production is inhibited offers the opportunity of evaluating the role of the gas in regulating several plant developmental processes such as leaf senescence, abscission, ripening and response to pathogens, and raises the possibility of being able to manipulate these genetically to improve the quality, storage life and nutritional value of many plants and plant products. □

Cis-trans recognition and subunit-specific degradation of short-lived proteins

Erica S. Johnson, David K. Gonda
& Alexander Varshavsky*

Department of Biology, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

THE N-end rule, a code that relates the metabolic stability of a protein to the identity of its amino-terminal residue¹, is universal in that different versions of the N-end rule operate in mammals²⁻⁵, yeast^{1,6,7} and bacteria (unpublished data). The N-end rule-based degradation signal comprises a destabilizing amino-terminal residue and a specific internal lysine residue^{1,6,7}. We now show that, in a multisubunit protein, these two determinants can be located on different subunits and still target the protein for destruction. Moreover, in this case (*trans* recognition) only the subunit that bears the lysine determinant is actually degraded. Thus an oligomeric protein can contain both short-lived and long-lived subunits. These insights have functional and practical implications.

In eukaryotes, ubiquitin-X- β -galactosidase (Ub-X- β gal) fusion proteins are deubiquitinated by an endogenous ubiquitin-specific protease to yield X- β gal test proteins^{1,5-7}. Depending on the identity of an amino-terminal residue X, an X- β gal is either long-lived or metabolically unstable^{1,5}. This amino-terminal degradation signal is manifested as the N-end rule.

The second determinant of the N-end rule-based degradation signal was discovered in experiments in which the first 38 residues of X- β gal were transplanted onto the amino terminus of a different test protein, dihydrofolate reductase (DHFR)⁶. The half-life ($t_{1/2}$) of the modified DHFR depended on the identity of its amino-terminal residue, provided, however, that at least one of the two lysines (Lys 15 or Lys 17) was present in the X- β gal-derived extension (Fig. 1a). We now show that this extension has the same role in the N-end rule-based degradation signal of X- β gal as it does when present in the X-DHFR proteins. To simplify the discussion, a β -galactosidase (β gal) polypeptide chain bearing the amino-terminal residue X and lysine (K in the single-letter amino-acid notation) at positions 15 and 17 (Fig. 1a) is denoted as XK. When both of these lysines are replaced with non-ubiquitinatable (arginine) residues, the resulting X- β gal is denoted as X Δ K.

When ³⁵S-labelled RK β gal (hereafter called RK), which bears a destabilizing amino-terminal residue, arginine (single-letter notation R), was incubated in ATP-supplemented rabbit reticulocyte extract, it was degraded by an ATP- and ubiquitin-dependent pathway (the N-end rule pathway) that operates in this *in vitro* system^{2,4,5} (Fig. 1b, curve 1). (Unless stated otherwise, the ATP supplementation of reticulocyte extract is implied.) Replacement of Lys 15/17 in RK ($t_{1/2}$ of ~1.6 h) with Arg residues yielded R Δ K β gal (R Δ K), a much longer-lived protein in reticulocyte extract ($t_{1/2}$ of ~11.5 h) (Fig. 1b, curves 1 and 2; see ref. 5 and Fig. 1 legend for the derivation of $t_{1/2}$). In agreement with previous evidence^{1,5,7}, the longer-lived R Δ K was ubiquitinated⁸ to a much lesser extent than was RK (Fig. 1c, lanes 2 and 4). Ubiquitin moieties in the multiply ubiquitinated RK occur as a multi-ubiquitin chain⁷ joined to either Lys 15 or Lys 17 of RK. The absence of these lysines (Fig. 1a) from R Δ K accounts for its much weaker ubiquitination (Fig. 1c, lanes 2 and 4). The above results directly support the earlier conjecture^{6,7} that Lys 15/17, being the main ubiquitination sites in XK, are, functionally, the principal second determinants of its degradation signal. However, these lysines are not the only

Received 12 March; accepted 1 May 1990.

1. Abeles, F. B. *Ethylene in Plant Biology* (Academic, New York, 1973).
2. Grierson, D. et al. *Phil. Trans. R. Soc. B* **334**, 399-410 (1986).
3. Davies, K. M., Hobson, G. E. & Grierson, D. *Pl. Cell Environ.* **11**, 729-738 (1988).
4. Davies, K. M. & Grierson, D. *Planta* **179**, 73-80 (1989).
5. Smith, C. J. S., Slater, A. & Grierson, D. *Planta* **168**, 94-100 (1986).
6. Slater, A. et al. *Plant molec. Biol.* **5**, 137-147 (1985).
7. Holdsworth, M. J., Schuch, W. & Grierson, D. *Plant molec. Biol.* **11**, 81-88 (1988).
8. Holdsworth, M. J. et al. *Nucleic Acids Res.* **15**, 731-739 (1987).
9. Holdsworth, M. J., Schuch, W. & Grierson, D. *Nucleic Acids Res.* **15**, 10600 (1987).
10. van der Krol, A. R. et al. *Nature* **333**, 866-869 (1988).
11. Smith, C. J. S. et al. *Nature* **334**, 724-726 (1988).
12. Sheehy, R. E., Kramer, M. & Hiatt, W. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8805-8809 (1988).
13. Pietrzak, M., Shilliter, R. D., Hohn, T. & Potrykus, I. *Nucleic Acids Res.* **14**, 5857-5868 (1986).
14. Bevan, M. W. *Nucleic Acids Res.* **12**, 8711-8722 (1984).
15. Bird, C. R. et al. *Plant molec. Biol.* **11**, 651-662 (1988).
16. Yang, S. F. *Current Topics in Plant Biochemistry and Physiology* Vol. 4 (eds Randall, D. D., Blevins, D. G. & Lasso, R. L.) 126-138 (University of Missouri, Columbia, 1985).
17. Sato, T. & Theologis, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6621-6625 (1989).
18. Raeder, V. & Broda, P. *Lett. appl. Microbiol.* **1**, 17-20 (1985).
19. Chomczynski, P. & Qasba, P. K. *Biochem. biophys. Res. Commun.* **122**, 340-344 (1984).
20. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1989).
21. Murray, M. G. & Thompson, W. F. *Nucleic Acids Res.* **8**, 4321-4325 (1980).
22. Grierson, D. & Tucker, G. A. *Plant* **157**, 174-179 (1983).
23. Hoffman, N. E. & Yang, S. F. *Pl. Physiol.* **69**, 317-322 (1982).

ACKNOWLEDGEMENTS. We thank Dr T. Solomos for advice about the ACC-oxidase assay. This work was supported by the AFRC, the Gatsby Charitable Foundation and an SERC Biotechnology Directorate Studentship to A.H. and was carried out under MAFF licences.

* To whom correspondence should be addressed.

detectable second determinants in RK, as indicated by incomplete metabolic stabilization of RΔK (relative, for instance, to VΔK; Fig. 1b, curves 2 and 4), and by the presence of ubiquitinated RΔK (Fig. 1c, lane 4). As RΔK lacks Lys 15/17, its ubiquitinated derivatives must have ubiquitin attached to one or more of the 20 lysines⁷ located throughout the rest of RΔK.

The degradation in reticulocyte extract of VKβgal (VK), which bears a stabilizing amino-terminal residue, valine (Val, V), although extremely slow, was also mediated by the N-end rule pathway, as it was reproducibly further reduced by the replacement of Lys 15 and 17 with non-ubiquitinatable (arginine) residues, yielding VΔK (3% and 1% degradation in 2 h for VK and VΔK, respectively; Fig. 1b, curves 3 and 4). (The degradation rate of any X-βgal in ATP-depleted reticulocyte extract is 0.2–0.4% degradation in 2 h (ref. 5 and data not shown).) As expected^{1,5}, the low level of ubiquitination of VK (Fig. 1c, lane 6) was further reduced for VΔK (Fig. 1c, lane 8). These results underscore the heuristic nature of the distinction between stabilizing and destabilizing residues in the N-end rule⁵.

In a monomeric substrate of the N-end-rule pathway, both

determinants of the degradation signal are located, by definition, within the same polypeptide chain, and are therefore said to be recognized in *cis* (Fig. 2a). However, for a multisubunit protein such as an X-βgal tetramer, one could also consider the possibility of *trans* recognition, in which the first and second determinants of the degradation signal reside within different subunits of X-βgal (Fig. 2b). (To simplify the discussion, the *cis* and *trans* concepts are illustrated in Fig. 2 for a dimeric protein.) To follow the fates of individual X-βgal subunits in an X-βgal tetramer, we prepared radioactively labelled Ub-X-βgal heterotetramers, the bulk of which contained one labelled and three unlabelled subunits per tetramer (Fig. 3a and its legend).

The reconstituted VK*-VK βgal tetramer (the ³⁵S-labelled subunit is designated by an asterisk) was degraded in reticulocyte extract at a slightly higher rate than the initial VK tetramer that had not been subjected to the dissociation-reassociation treatment (*t*_{1/2} of ~20 h and ~50 h, respectively; Fig. 1b, curve 3; Fig. 3b, curve 3). This degradation, although ATP-dependent (data not shown), was not mediated by the N-end rule pathway because the labelled subunit in a recon-

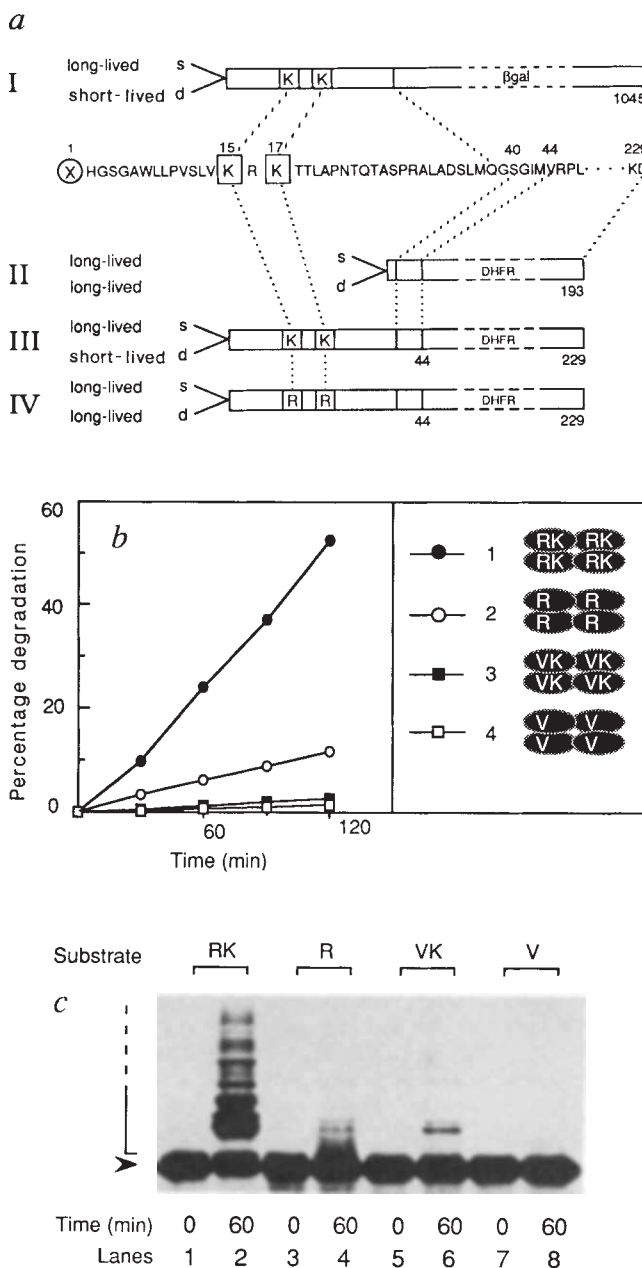


FIG. 1 The second determinant of the N-end rule-based degradation signal. **a**, Adding the second determinant to dihydrofolate reductase (DHFR). The constructs shown were expressed as ubiquitin fusions whose deubiquitination yielded the desired test proteins^{5,6}. I, X-βgal bearing either a stabilizing (s) or a destabilizing (d) amino-terminal residue. The *t*_{1/2} of an s-βgal in *S. cerevisiae* at 30 °C is more than 20 h. The *t*_{1/2} of a d-βgal under the same conditions varies from ~30 min for Ile-βgal to ~2 min for Arg-βgal (refs. 1,6). The (unshaded) amino-terminal region of X-βgal was derived in part from an internal sequence of the *lac* repressor⁶, and is also shown in single-letter amino-acid notation. II, unmodified X-DHFR, whose *t*_{1/2} in *S. cerevisiae* differs at most fivefold between an s-DHFR and a d-DHFR (ref. 6). III, same as II but with the 38-residue amino-terminal region of X-βgal positioned upstream of the original amino terminus of X-DHFR. The *t*_{1/2} of these extension-bearing X-DHFR proteins differs up to ~500-fold between an s-DHFR and a d-DHFR (ref. 6). IV, same as III but with Lys 15/17 replaced with Arg residues (see main text and ref. 6). **b**, Kinetics of degradation, in ATP-supplemented reticulocyte extract, of ³⁵S-labelled X-βgals bearing either a stabilizing (V) or a destabilizing (R) amino-terminal residue, and either containing or lacking the second determinant of the degradation signal. In the sketches of βgal tetramers on the right, the left-hand letter within each subunit identifies its amino-terminal residue (R or V). The presence of Lys 15/17 (see **a**) is denoted by the letter K (Lys). **c**, Ubiquitination of X-βgals in reticulocyte extract. Equal amounts of [³⁵S]X-βgals shown in **b** were incubated in ATP-supplemented reticulocyte extract for either 0 or 60 min at 37 °C, followed by SDS-PAGE and fluorography. An arrowhead denotes X-βgal. The half-open square bracket indicates the bands of ubiquitinated X-βgal.

METHODS. The pKK233-2-based *Escherichia coli* expression vectors pKKUb-Arg-βgal and pKKUb-Val-βgal encoding, respectively, Ub-RKβgal and Ub-VKβgal, have been described⁵. The constructs encoding Ub-RΔKβgal and Ub-VΔKβgal were made using oligonucleotide-directed mutagenesis²². Owing to the construction route taken, Leu-Ala at the positions 39–40 of the original VK and RK βgals was replaced, in VΔK and RΔK, by Gly-Ser. These changes were confined to the *E. coli lacI*-encoded 45-residue extension present at the amino termini of our X-βgal test proteins⁵. To construct pKKUb-Arg-ΔLys-βgal, the large *SalI*–*BamHI* fragment of pKKUb-Val-ΔLys-βgal was ligated to the small *SalI*–*BamHI* fragment of pKKUb-Arg-βgal. Detailed construction protocols are available on request. Ub-X-βgal proteins were expressed in *E. coli*, metabolically labelled with [³⁵S]methionine, and purified by affinity chromatography as described⁵, except that 3 mCi of ³⁵S-Translabel (ICN) were used for labelling, and that MgCl₂ was omitted from the final storage buffer. The specific radioactivity of Ub-X-βgal was 6 × 10⁵–8 × 10⁵ c.p.m. μg⁻¹. Rabbit reticulocyte extract was prepared and assayed for the degradation of X-βgals as previously described⁵, including a 10-min preincubation in ATP-depleted extract⁵ to deubiquitinate Ub-X-βgal. The degradation of [³⁵S]X-βgals was assayed by measuring acid-soluble radioactivity⁵. SDS-PAGE was carried out in 6.5% polyacrylamide, 0.18% bisacrylamide gels, with subsequent fluorography. The ATP- and ubiquitin-dependent degradation of X-βgal in reticulocyte extract obeyed first-order kinetics for at least the first 2 h, making it possible to compare the degradation of different X-βgal proteins by comparing their half-lives in the extract⁵. The *t*_{1/2} of RKβgal did not change when its initial concentration in the extract was reduced 20-fold (data not shown).

stituted ΔK^* - ΔK (all of whose subunits lacked both determinants of the degradation signal) was degraded in reticulocyte extract at a similar rate (Fig. 3b, curves 3 and 6). This additional slow degradation was therefore the background of our assays with reconstituted X- β gal tetramers.

When the labelled VK subunit was present in the VK^* -RK heterotetramer, it was multiply ubiquitinated and degraded in reticulocyte extract with a $t_{1/2}$ of ~ 3.3 h (Fig. 3b, curve 1; Fig. 3d, lane 1). In contrast, when the labelled VK subunit was part of the VK^* -VK tetramer, its $t_{1/2}$ was 20 h, and it showed very little ubiquitination (Fig. 3b, curve 3; Fig. 3d, lane 3). In a control experiment, the labelled VK homotetramer was mixed with an excess of the unlabelled RK homotetramer, and incubated in reticulocyte extract without the dissociation-reassociation pretreatment. In these conditions, VK was long-lived, with a $t_{1/2}$ of ~ 50 h (data not shown), identical to the $t_{1/2}$ of VK in the absence of RK (Fig. 1b, curve 3). These results indicated that a subunit bearing a stabilizing amino-terminal residue can be metabolically destabilized if it is physically associated with subunits containing the N-end rule-based degradation signal. These results also suggested the existence of *trans* recognition in the N-end rule pathway (Fig. 2b). In other words, these results suggested that recognition of a destabilizing amino-terminal

residue in an RK subunit of the VK^* -RK tetramer could promote multi-ubiquitination of the second-determinant lysine (Lys 15 or Lys 17) in the VK subunit of the same heterotetramer, resulting in the degradation of this otherwise long-lived subunit.

Remarkably, the metabolic instability of the VK subunit in the VK^* -RK heterotetramer did not depend on the degradation of the tetramer's intrinsically short-lived RK subunits: the VK subunit had a $t_{1/2}$ of ~ 3.3 h in reticulocyte extract whether it was part of either VK^* -RK or VK^* - ΔK (Fig. 3b, curves 1 and 2). The ΔK subunits of the latter heterotetramer lacked the second-determinant Lys 15/17 (Fig. 1a) and were therefore much longer-lived in reticulocyte extract than the otherwise identical RK subunits (Fig. 1b, curves 1 and 2). These results (Fig. 3b, curves 1-3; Fig. 3d, lanes 1-3) directly showed the existence of *trans* recognition of short-lived subunits within X- β gal tetramers. An additional control experiment (see the Fig. 3 legend) confirmed that the *trans* recognition-dependent degradation of the VK subunit is mediated by the N-end rule pathway.

Although the labelled VK subunit was short-lived within VK^* -RK, the otherwise identical ΔK subunit lacking the second-determinant Lys 15/17 (Fig. 1a) was much more stable in the same tetramer background (Fig. 3b, curves 1 and 4). One implication of this result is that the *trans* recognition-mediated degradation of a subunit bearing a stabilizing amino-terminal residue depends on the presence of the second determinant (multi-ubiquitination site) in the *trans*-targeted subunit. In addition, this result indicated that an oligomeric protein can contain both short-lived and long-lived subunits. Indeed, although the ΔK subunit of ΔK^* -RK was long-lived in reticulocyte extract ($t_{1/2}$ of ~ 11.5 h; Fig. 3b, curve 4), the RK subunits of this heterotetramer were much more unstable, inasmuch as the RK subunit of RK^* - ΔK had a $t_{1/2}$ of ~ 2.0 h (Fig. 3c, curve 10). The ΔK subunit was degraded slightly faster in ΔK^* -RK than in ΔK^* - ΔK (Fig. 3b, curves 4 and 5). It is not known whether this small effect is due to an incomplete subunit specificity of degradation, or whether release of the intact ΔK subunit (during the degradation of RK within ΔK^* -RK) makes this subunit a substrate for other proteases in reticulocyte extract.

Another implication of these results is that the degradation of short-lived subunits is not inhibited by the presence of long-lived subunits in the same oligomeric protein. Specifically, the $t_{1/2}$ of 2.0 h for the RK subunit in reticulocyte extract was the same irrespective of whether this subunit was part of a homotetramer (Fig. 3c, curve 7) or was surrounded by the long-lived ΔK subunits (Fig. 3c, curve 10).

The differential degradation of individual subunits within X- β gal heterotetramers was not caused by a rapid dissociation-reassociation of X- β gal subunits in reticulocyte extract that might allow short-lived subunits to be targeted and degraded while in a transient monomeric state. The β -galactosidase tetramer is very stable, requiring strongly denaturing solvents for its dissociation⁹⁻¹¹. Furthermore, when labelled VK homotetramers and an excess of unlabelled ΔK homotetramers were incubated together in reticulocyte extract, the VK subunits were long-lived ($t_{1/2}$ of ~ 50 h; data not shown), in contrast to the metabolic instability of the VK subunit in the VK^* - ΔK heterotetramer ($t_{1/2}$ of ~ 3.3 h; Fig. 3b, curve 2). We conclude that the degradation of an oligomeric protein by the N-end rule pathway is subunit-specific.

XK β gal subunits bearing a stabilizing amino-terminal residue can be targeted for degradation only in *trans* (see above and Fig. 2b). On the other hand, the RK subunit, which contains both determinants of the degradation signal, can be targeted either in *cis* or in *trans* within RK^* -RK and RK^* - ΔK , but only in *cis* within RK^* -VK and RK^* - ΔK (Figs 2 and 3c). Nonetheless, the degradation time courses (as well as ubiquitination levels) of the RK subunits in reticulocyte extract were indistinguishable for all these tetramer backgrounds (Fig. 3c, curves 7-10; Fig. 3d, lanes 7-10). Thus, the possibility of *trans* recognition does not increase the rate of degradation of an

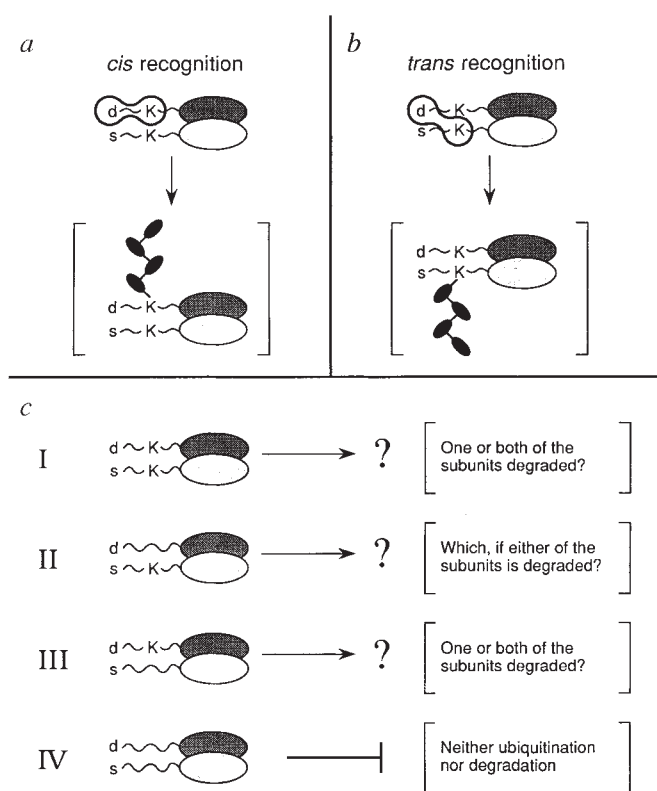
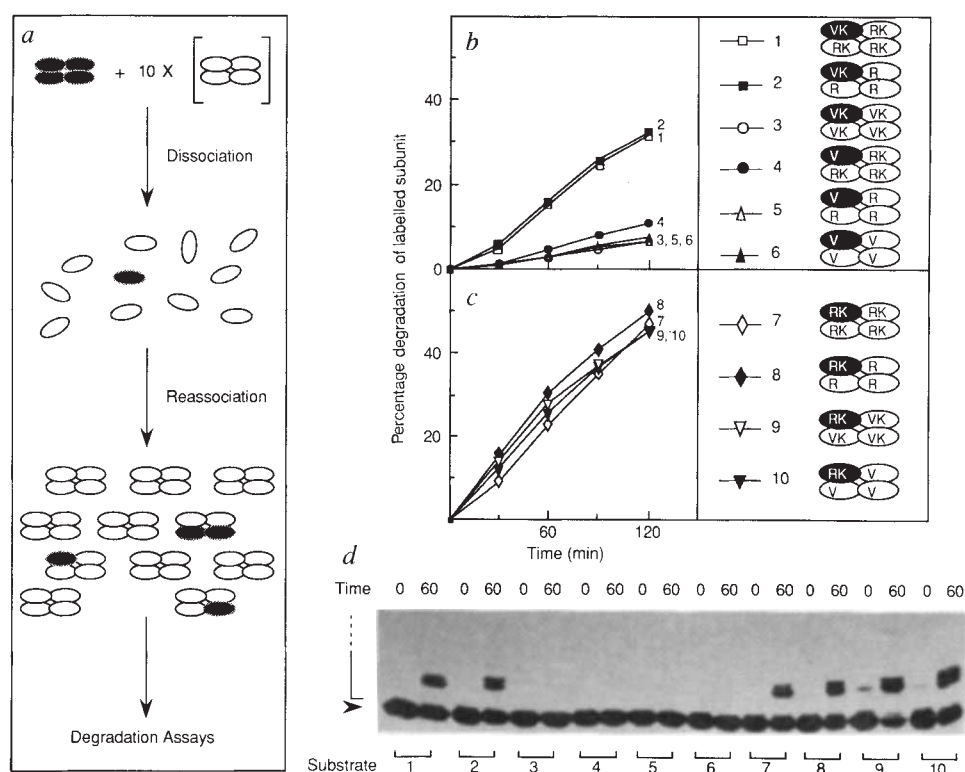


FIG. 2 *Cis* and *trans* recognition in the N-end rule pathway. **a**, *Cis* recognition of the complete degradation signal present in one subunit of a dimeric substrate. The other subunit of the dimer lacks the first of the two determinants of the signal. The letters d and s denote, respectively, a destabilizing and a stabilizing amino-terminal residue, with the former being the first determinant of the complete signal^{1,5-7}. The letter K denotes the second-determinant lysine residue⁶. Black ovals denote a multi-ubiquitin chain⁷ joined to the targeted second-determinant lysine. **b**, *Trans* recognition, in which the first (d) and second (K) determinants of the degradation signal are targeted while residing in different subunits of a dimeric substrate. **c**, Different arrangements of the first and second determinants between the subunits. The questions are answered in the text. For structure IV, which lacks the second-determinant lysine(s), the answer is expected from previous evidence⁶ (see also Fig. 1a).

FIG. 3 ATP-dependent ubiquitination and degradation of X- β -gal heterotetramers in reticulocyte extract. **a**, Design of procedure for reconstituting 35 S-labelled X- β -gal heterotetramers most of which contain one labelled and three unlabelled subunits per tetramer. Black and white ovals represent 35 S-labelled and unlabelled subunits, respectively. **b**, **c**, Kinetics of degradation, in ATP-supplemented reticulocyte extract, of 35 S-labelled X- β -gals produced as described in **a**. In the sketches of β -gal tetramers on the right, the designations are as described in the Fig. 1b legend. **d**, Ubiquitination, in ATP-supplemented reticulocyte extract, of 35 S-labelled subunits of X- β -gals shown in **b** and **c**. Equal amounts of these proteins were incubated in ATP-supplemented reticulocyte extract for either 0 or 60 min at 37 °C, followed by SDS-PAGE and fluorography. The pairs of time points are identified with the sample numbers assigned in **b** and **c**. For other designations, see the Fig. 1c legend.

METHODS. Variants of Ub-X- β -gal subunits were constructed, prepared (as homotetramers), and purified as described in the Fig. 1 legend. To prepare heterotetramers²³ of Ub-X- β -gal, a mixture of 18 μ g of [35 S]Ub-X- β -gal and 180 μ g of unlabelled Ub-X- β -gal (whose amino-acid sequence was either identical to or different from that of the labelled Ub-X- β -gal), at a total protein concentration of 1.5–3 mg ml⁻¹ in storage buffer (50% (v/v) glycerol, 0.1 mM Na-EDTA, 1 mM DTT, 40 mM Tris-HCl (pH 7.5)) was diluted to a protein concentration of 80 μ g ml⁻¹ with buffer A (10 mM Na-EDTA, 5 mM DTT, 20 mM Na-HEPES (pH 7.2)) containing 2 M NaCl and 8 M (deionized) urea. After 1 h at 0 °C, the solution was dialysed against changes of buffer A containing, successively, 8 M urea, 4 M urea and 10 mM NaCl, 2 M urea and 10 mM NaCl, 1 M urea and 10 mM NaCl, and finally, against two changes of buffer A plus 10 mM NaCl. Dialysis was carried out with 10 Ub-X- β -gal samples at a time, at 4 °C, in a Multiple Dialyzer (Spectrum) for ~14 h per change of buffer. Each of the dialysed samples was brought to 1.6 M in NaCl and 10 mM in MgCl₂, added to ~0.4 ml APTG-agarose⁵ and rocked overnight at 4 °C. The suspension was then loaded into columns, which were processed for affinity purification of Ub-X- β -gal as described⁵. At a 1:10 ratio of labelled to unlabelled Ub-X- β -gal subunits, random reassociation should result in 75% of the labelled subunits residing in tetramers containing one labelled subunit per tetramer, with most of the rest of the labelled subunits (23%) present at two subunits per tetramer. In a test of this prediction, a 1:10 mixture of two β -gal-based homotetramers containing amino-terminal extensions of different sizes was subjected to the dissociation-reassociation



ation protocol. Electrophoresis of the resulting tetramers in a nondenaturing gel produced the expected distribution of heterotetramer compositions (data not shown). The enzymatic activity of the reconstituted Ub-X- β -gal proteins (determined as described²⁴) varied between different preparations from 50 to 75% of the activity of their initial (homotetramer) counterparts. Degradation assays in reticulocyte extract and SDS-PAGE were carried out as described in the Fig. 1 legend. Less than 0.4% of either the initial or reconstituted X- β -gal tetramers were degraded in 2 h in ATP-depleted reticulocyte extract (data not shown). An additional control experiment (not shown) has exploited the existence of secondary destabilizing residues in the N-end rule^{4,5,25}. Specifically, we found that the labelled VK subunit was short-lived within the VK*-E Δ K β -gal tetramer in reticulocyte extract (E Δ K bears Asp (E), a secondary destabilizing amino-terminal residue^{4,5,25}). However, when the same assay was carried in a tRNA-depleted reticulocyte extract (this treatment converts secondary and tertiary destabilizing residues into stabilizing ones^{4,5,25}), the VK subunit became long-lived in VK*-E Δ K but not in VK*-R Δ K (the latter subunit bears a primary destabilizing amino-terminal residue⁵). These data (not shown) directly confirmed that the *trans* recognition-dependent degradation of the VK subunits is mediated by the N-end rule pathway.

X- β -gal subunit if the *cis* recognition mode is available as well. At the same time, for an X- β -gal, *trans* recognition alone is comparable in efficiency to *cis* recognition alone, because the VK subunit (targetable only in *trans*) was degraded in either VK*-RK or VK*-R Δ K at ~65% of the rate observed for the RK subunits which could be targeted at least in *cis* ($t_{1/2}$ of ~3.3 h and ~2.0 h, respectively; Fig. 3b, curves 1 and 2; Fig. 3c, curves 7–10). The lack of influence of *trans* recognition in the presence of the *cis* recognition option is not due to inhibition of the *trans* mode of recognition by the *cis* mode, inasmuch as the VK subunit was degraded at indistinguishable rates in either VK*-RK or VK*-R Δ K (Fig. 3b, curves 1 and 2). A likely explanation of these results is that the *cis* recognition alone may be sufficient for the degradation of an X- β -gal subunit to proceed at a rate limited by some other step in the N-end rule pathway.

The existence of *trans* targeting in the N-end rule pathway (Figs 2 and 3) is consistent with the previously proposed model⁶ in which an N-end-recognizing (E3) protein (either alone or

complexed with a specific ubiquitin-conjugating (E2) enzyme) has a binding site for a substrate's destabilizing amino-terminal residue as well as a lysine-binding site. Occupation of both of these sites by a proteolytic substrate is postulated to be required for multi-ubiquitination to commence at the bound lysine of the substrate⁶. If the two binding sites of the recognition complex are spatially fixed relative to one another, binding of substrate would require either a specific spatial arrangement of the substrate's two determinants or mobility of these determinants relative to one another, so that a correct spatial arrangement is achieved transiently, but frequently enough for both of them to be bound by the recognition complex.

Segmental mobility of the determinant-containing region (or regions) can account for both *cis* and *trans* recognition in a short-lived X- β -gal. Indeed, the non- β -gal extension at the amino termini of the X- β -gal proteins (Fig. 1a), being derived from an internal region of the *lac* repressor, is likely to be disordered (segmentally mobile) when present in an unnatural (amino-

terminal) location in an unrelated protein such as β -galactosidase⁶. Because of segmental mobility of the extension, its Lys 15/17 residues (Fig. 1a) could occur transiently in spatial proximity to a destabilizing amino-terminal residue of either their own or a different subunit in the same X- β gal tetramer, the latter allowing *trans* recognition.

We have also shown (Fig. 3 and discussion above) that, in a metabolically unstable X- β gal tetramer, only those subunits containing the second (but not necessarily the first) determinant of the degradation signal are actually degraded. In other words, degradation is confined to those subunits that can be ubiquitinated. (Although some earlier data on protein turnover (see refs 12–14) were consistent with the notion of subunit-specific degradation, none of these studies directly demonstrated it.) What mechanism could underlie this aspect of protein turnover? In one class of models, the selective (and apparently processive⁷) degradation of a subunit bearing a multi-ubiquitin chain would be temporally coupled to its dissociation from adjacent subunits in an oligomeric substrate. For this to occur, the ATP-dependent, ubiquitin conjugate-degrading protease⁸ would recognize the multi-ubiquitin chain⁷, use it to identify the subunit to destroy, and thereafter track along the subunit's polypeptide chain, unfolding it before degradation. *A priori*, the dissociation and degradation need not be temporally coupled: a mechanochemical process analogous to the one above might be used to dissociate the subunit containing the multi-ubiquitin chain from an oligomeric substrate, with the degradation of the subunit following rather than accompanying its dissociation.

Subunit specificity is likely to be a general feature of selective protein degradation because alteration of the functions of protein complexes by combinatorial variation of their subunit composition is a common theme in biological regulation^{15–20}. The generality of this new aspect of protein turnover is supported

by the recent finding²¹ that subunit-specific degradation is also characteristic of at least one of the two degradation signals present in the naturally short-lived transcriptional repressor MAT α 2 of the yeast *Saccharomyces cerevisiae*. Neither of these signals operates in the N-end rule pathway²¹.

Potential applications of *trans* recognition include the ability to destabilize metabolically any otherwise long-lived protein by targeting it for degradation in *trans*. The feasibility of these applications is being ascertained. □

Received 18 January; accepted 10 May 1990.

1. Bachmair, A., Finley, D. & Varshavsky, A. *Science* **234**, 179–186 (1986).
2. Varshavsky, A. *et al.* in *Ubiquitin* (ed. Rechsteiner, M.) 287–324 (Plenum, New York, 1988).
3. Townsend, A. *et al.* *J. exp. Med.* **168**, 1211–1224 (1988).
4. Reiss, Y., Kaim, D. & Herskko, A. *J. Biol. Chem.* **263**, 2693–2698 (1988).
5. Gonda, D. K. *et al.* *J. Biol. Chem.* **264**, 16700–16712 (1989).
6. Bachmair, A. & Varshavsky, A. *Cell* **56**, 1019–1032 (1989).
7. Chau, V. *et al.* *Science* **243**, 1576–1583 (1989).
8. Herskko, A. *J. Biol. Chem.* **263**, 15237–15240 (1988).
9. Zipser, D. *J. molec. Biol.* **7**, 113–121 (1963).
10. Shifrin, S. & Steers, E. Jr *Biochim. biophys. Acta* **133**, 463–471 (1967).
11. Givol, D., Craven, G. R., Steers, E. & Anfinsen, C. B. *Biochim. biophys. Acta* **113**, 120–125 (1966).
12. Schimke, R. T. *Adv. Enzymol.* **37**, 135–187 (1973).
13. Goldberg, A. K. & Dice, J. F. *A. Rev. Biochem.* **43**, 835–869 (1974).
14. Hare, J. F. & Hodges, R. *J. Biol. Chem.* **257**, 3575–3580 (1982).
15. Evans, T. *et al.* *Cell* **33**, 389–396 (1983).
16. Murray, A. W., Solomon, M. J. & Kirschner, M. W. *Nature* **339**, 280–286 (1989).
17. Abel, T. & Maniatis, T. *Nature* **341**, 24–25 (1989).
18. Goutte, C. & Johnson, A. D. *Cell* **52**, 875–882 (1988).
19. Ingham, P. W. *Nature* **335**, 25–34 (1988).
20. Scott, M. P. & Carroll, S. B. *Cell* **51**, 689–698 (1987).
21. Hochstrasser, M. & Varshavsky, A. *Cell* **61**, 697–708 (1990).
22. Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (Wiley, New York, 1987).
23. Ullman, A. & Monod, J. *Biochem. biophys. Res. Commun.* **35**, 35–42 (1969).
24. Guarente, L. *Meth. Enzym.* **101**, 181–183 (1983).
25. Ferber, S. & Ciechanover, A. *Nature* **326**, 808–811 (1987).

ACKNOWLEDGEMENTS. We thank members of our laboratory, especially B. Bartel, J. Dohmen, M. Hochstrasser, K. Madura, I. Ota and J. Tobias for comments on the manuscript, and B. Doran for secretarial assistance. This work was supported by the NIH (A.V.), E.S.J. was an NSF Predoctoral Fellow. D.K.G. was a Fellow of Jane Coffin Childs Memorial Fund for Medical Research.

Striking conservation of TFIID in *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*

John D. Fikes, Daniel M. Becker, Fred Winston* & Leonard Guarente

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

* Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

EUKARYOTIC promoters contain binding sites for basic transcription factors and gene-specific activator proteins^{1–6}. The transcription factors interact at the TATA box, which lies close to the position of transcription initiation. Activators typically bind to distant sites that can lie kilobases away from the initiation site. The factor TFIID binds specifically to the TATA box to initiate an ordered pathway of assembly of the basic transcription factors^{5,7}. Biochemical analyses have shown that human and *Saccharomyces cerevisiae* TFIID are functionally interchangeable *in vitro*^{8–10}. To study further the functional conservation of this critical factor, we are surveying proteins from divergent organisms that can substitute *in vivo* for the *S. cerevisiae* TFIID. We report here the isolation of a unique gene from *Schizosaccharomyces pombe* that fully complements a null mutation in *SPT15*, the gene that encodes TFIID^{11–15} in *S. cerevisiae*. The *Schiz. pombe* gene encodes a protein 93% identical (166/178) to *S. cerevisiae* TFIID in a region consisting of a direct repeat.

A comprehensive complementary DNA library was generated from *Schiz. pombe* poly(A)⁺ RNA by methods to be described

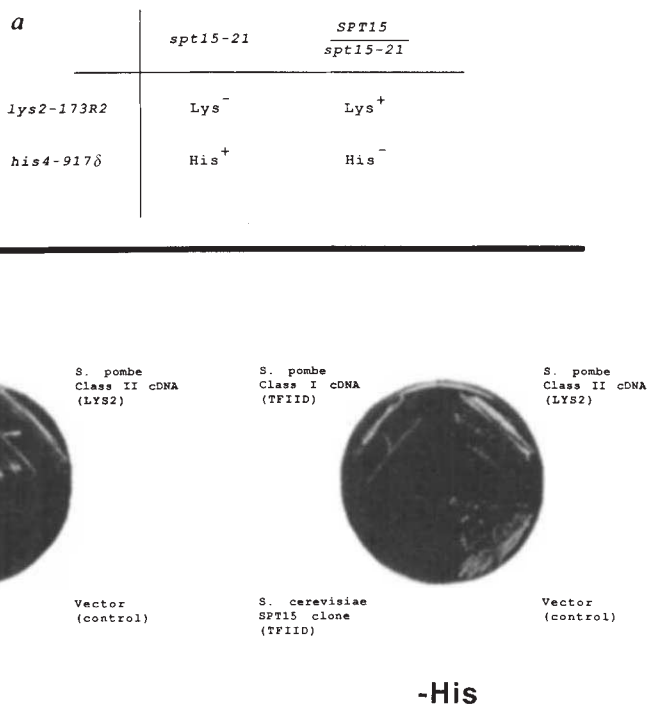
in detail elsewhere. The cDNA was inserted adjacent to a strong constitutive *S. cerevisiae* promoter, the alcohol dehydrogenase-1 (ADH1) promoter¹⁶, in a 2- μ m URA3 plasmid. Following electroporation, plasmid DNA from 2.2×10^7 primary *Escherichia coli* transformants was prepared. The size distribution of cDNA inserts approximated that of the messenger RNA (data not shown), and 12/12 randomly chosen plasmids were found to contain cDNA inserts.

SPT15 mutations affect transcription of *S. cerevisiae* genes that bear Ty insertions in their promoters¹⁷. To isolate clones that complement *spt15* mutations, we used *S. cerevisiae* strain FW1259, which contains the *spt15-21* mutation in TFIID and insertions of portions of the Ty element at the *LYS2* and *HIS4* loci¹². As shown in Fig. 1a, the phenotype of FW1259 (Lys[−]His⁺) can be converted to Lys⁺His[−] by introduction of an *SPT15* gene *in trans*. We anticipated that introduction of cDNA clones encoding *Schiz. pombe* TFIID activity would cause a similar phenotypic conversion in this strain.

FW1259 was transformed with library DNA and Ura⁺Lys⁺ colonies were selected. Approximately 1 in 10^4 yeast transformants was converted to Lys⁺. Both Lys⁺His[−] (class I) and Lys⁺His⁺ (class II) phenotypes were observed (Fig. 1b). To verify that these phenotypic alterations were due to cloned cDNA and not to strain reversion, we recovered plasmid from transformants of each class and retransformed strain FW1259. In every case, phenotypic conversion was conferred by the plasmid. We anticipated that a Lys⁺ phenotype might also be obtained through expression of the *Schiz. pombe* *LYS2* homologue. To exclude these clones from further analysis, we also transformed each plasmid into a Lys[−] strain (L3283) to identify clones that directly complement the *Lys2* defect. This analysis revealed that all of the class II clones and none of the class I clones bore the *Schiz pombe* *LYS2* homologue, encoded by a 4-kilobase (kb) cDNA.

FIG. 1 *Schiz. pombe* suppressors of the *spt15-21* mutation. *a*, As previously reported¹⁷, introduction of the *S. cerevisiae* *SPT15* gene, encoding wild-type TFIID, into *S. cerevisiae* strain FW1259 (*Mata his4-917 δ lys2-173R2 ura3-52 trp1 Δ 1 spt15-21*) converts the phenotype from Lys⁻ His⁺ to Lys⁺ His⁻. This phenotypic conversion was exploited to select for *Schiz. pombe* cDNA clones encoding TFIID function. *b*, Two classes of *Schiz. pombe* cDNA clones were obtained in the selection for Lys⁺ transformants. Representative transformants of each class are shown restreaked on plates lacking the indicated amino acid. Strains harbouring class I cDNA clones exhibit a Lys⁺ His⁻ phenotype similar to that conferred by *S. cerevisiae* *SPT15*, but distinguished by slightly greater growth on plates lacking histidine. Class II cDNA clones were found to confer a Lys⁺ His⁺ phenotype, a combination distinct from that conferred by *SPT15* or vector alone. In these transformants, a slight general growth defect is observed. Further analysis revealed that the class II clones encode the *Schiz. pombe* *LYS2* gene.

METHODS. A *Schiz. pombe* cDNA library was constructed using a plasmid vector that directs high-level expression in *S. cerevisiae*. Twenty micrograms of this URA3-marked library was transformed into *S. cerevisiae* strain FW1259 (ref. 23). An estimated 2×10^6 Ura⁺ transformants were plated on minimal media lacking uracil and lysine, and Ura⁺ Lys⁺ colonies were obtained at a frequency of about 1 in 10^4 . Plasmids were recovered from selected transformants and retransformed into FW1259, confirming in each case linkage of the Lys⁺ phenotype to the plasmid. The His phenotype of these transformants allowed grouping

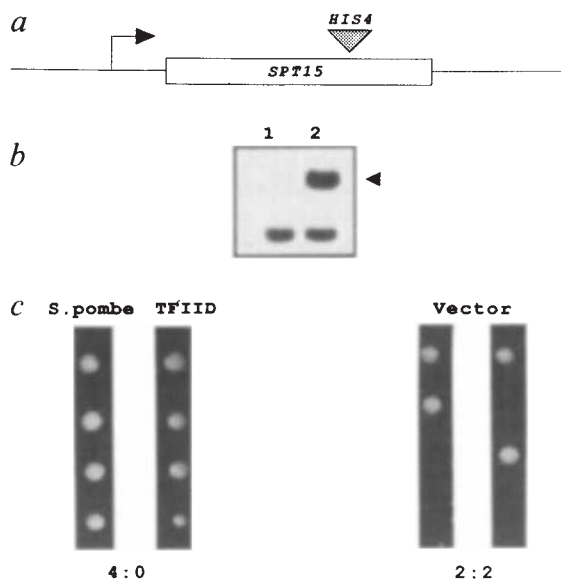


into two classes. Plasmids were also transformed into the Lys2⁻ *S. cerevisiae* strain L3283 (*Mata ura3-52 lys2-201 his3-200*; G. Fink, personal communication). Complementation of the *lys2* defect in this strain identified *Schiz. pombe* clones encoding the *Lys2* homologue.

To determine whether the class I *Schiz. pombe* clones could substitute for all the activities of the *S. cerevisiae* TFIID, we tested whether these clones could rescue a null mutation in the *SPT15* gene. An *SPT15* insertion mutation was constructed in a diploid strain (Fig. 2*a, b*), and the resulting heterozygous strain (*SPT15/spt15::HIS4*) was transformed with either of two different class I clones, or with vector alone. Transformants were sporulated, and tetrads were analysed (Fig. 2*c*). The genetic data show that the class I *Schiz. pombe* cDNA clones fully complement a disruption of the *SPT15* gene encoding the *S. cerevisiae* TFIID (see Fig. 2 legend).

Eight class I clones bore similar 1.3-kb inserts; the sequence of one clone was determined, revealing a 231-codon open reading frame (Fig. 3). Comparison of the predicted amino-acid sequence with the *S. cerevisiae* TFIID sequence identifies only very weak homology at the amino-terminal end (Fig. 4*a*). In striking contrast, 166/178 amino acids at the carboxy terminus are identical between the two TFIID proteins. Furthermore, of the 12 divergent residues, 5 represent conservative changes. The degree of conservation in this 178-amino-acid region (93% identity) equals that observed for the most highly conserved proteins. For example, the conservation between histone proteins of *Schiz.*

FIG. 2 *Schiz. pombe* class I clones (TFIID) can fully complement a null mutation in the *S. cerevisiae* *SPT15* gene. *a*, The *SPT15* gene present in plasmid DE28-6¹² was disrupted by insertion of the *HIS4* gene into the unique *Bal*I restriction site. A 5.9-kb restriction fragment encompassing the *spt15::HIS4* construct was purified and used to transform a phenotypically His⁻ diploid strain (*Mata α his4-619/his4-619 ura3-52/ura3-52 leu2-3, 112/LEU2 lys2-801am/LYS2*) to His⁺, thereby disrupting one of the two chromosomal copies of *SPT15* by the method of Rothstein²⁴. *b*, Southern analysis confirmed replacement of one chromosomal copy of *SPT15* with *spt15::HIS4*. Genomic DNA prepared from a haploid strain, lane 1, and from six His⁺ diploid transformants, one of which is shown in lane 2, was probed with *SPT15*. An intact *SPT15* gene is detected in each lane as the lower (2.3 kb) *Bam*HI/*Pst*I restriction fragment. The additional band (5.7 kb) in lane 2, denoted by the arrow, identifies the disrupted allele. *c*, The heterozygous *SPT15/spt15::HIS4* strain was transformed with class I cDNA clones or with vector. These transformants were sporulated and tetrads from each were analysed using standard techniques²³. Representative tetrads, grouped vertically, are shown. Eight tetrads were analysed from transformants harbouring vector alone; all segregated two viable and two inviable spores. The viable spores were all His⁻, indicating presence of the wild-type *SPT15*⁺ allele. In contrast, four viable spores were obtained in nine tetrads from transformants harbouring the cDNA clones. These tetrads all contained two His⁺ (*spt15::HIS4*) and two His⁻ spores. Every segregant which carried the null allele of *SPT15* harboured the cDNA clone, as shown by its Ura⁺ phenotype. As expected, tetrads with fewer than four viable spores were also observed, presumably arising from plasmid loss or unequal plasmid segregation during sporulation.



```

1  ATTAGGGGTGGGATAACGCAATTCACGAGTAGCGTAAATAATTAATTGATTACTGCA
60  ATGGATTTCGCTTTACCCACCACGGCTCGCAAGGAGTGCCTTTATGAATAACTCTTCT
   M D F A L P T T A S Q A S A F M N N S S
120  TTAACGTTCCCTGTTCTCCCAATGCCAATAAGAGGTACAAATGACACGGCAGATTCC
   L T F P V L P N A N N E A T N E T A D S
180  GGGGATGCAGAAATTTCAAAAAATGAAGGTGATCTGGCATTGTTCAACCCCTTCAAAAT
   G D A E V S K N E G V S G I V P T L Q N
240  ATTGTTGCTACTGTAACTTAGACTGCTGCTGATCTCAAAACTATTGGCGTACATGCA
   I V A T V N L D C R L D L K T I A L H A
300  CGTAATGCAGAAATCAACCCAAACGTTTTCGGCTGTTATTATGCGTATCGTGAACCC
   R N A E Y N P K R F A A V I M R I R E P
360  AAGTCTACTGCATTGATTTTCGGCTGCGTAAATGGTGTGTTGGGTGGCAAAATCCGAG
   K S T A L I F A S G K M V V L G G K S E
420  GATGACTCCAAGCTCGCTCTAGCAAGTATGCGGTATCAACCAAACTCGGTTTAAAT
   D D S K L A S R K Y A R I I Q K L G F N
480  GCCAAGTTCACGGATTTAAGATTGCAACATTTAGGAAGTTCGATGTTAAATTTCCA
   A K F T D F K I Q N I V G S C D V K F P
540  ATTCGTTTGAAGGTTTGGCTTACTCCACGGTACTTTCTCATCTTATGAGCTGAGTTG
   I R L E G L A Y S H G T F S S Y E P E L
600  TTTCCCGGTTTGATTATCGCATGGTAAACCAAAAGTTGTTCTATTGATTTTGTTCF
   F P G L I Y R M V K P K V V L L I F V S
660  GGTAAATGTTTAACTGCTGCAAGTCCGTGAGGAATTTACCAAGCTTTTGAAGCC
   G K I V L T G A K V R E E I Y Q A F E A
720  ATTTATCCAGTATTGCTGAATTCGAAAACATTAA
   I Y P V L S E F R K H *
      GGCATGTCAACAGTTATCACACAG
780  TTTTGTGCAATGTTATGGGTATTGTTGGGTAAAGACCTTAGTCGGGTCAGGTATT
840  TTTTGATCTTTAACATTAAACCCCTTAAAGCTTTCGGGAGAAATCCCTTCACTGTAAGT
900  CATCATTATGAGTATTTTATATGATGACAGATGATAGATATCCTTTGAAAAGTTTTT
960  TCTTCAAAATCAGAACTGATCTGTGAGTTCTTCTTTTACGGGCGCAACAGCTTATT
1020  CAAGAAAATATAGAGGCTGAAATTCGCTATCTTATTGAAAACCTGCAGTTGATGAATG
1080  TGTCCTCTTAAAGTGGTGCAGTGTCTTCGGATACACTGTTTTCCTCTGTGTTTA
1140  TCATTAAACATTAGTTGAGCATTGCTTCAATAATCTTCTAGCTTCAGGATCGTGGCCA
1200  AATTGGTACAATGTTCTGTATAGCTTTGTATAATTCATTTTTCATGATCGTGAATAAT
1260  AATAATATCTTTTAAATGTAAGAAAAA

```

FIG. 3 Complete nucleotide sequence of a cDNA encoding *Schiz. pombe* TFIID. The entire 1,295-nucleotide cDNA sequence is shown, with translation of a 231-codon open reading frame. Representative in-frame stop codons are marked by asterisks.

METHODS. The cDNA was excised from a representative class I clone by cutting at the *NotI* sites flanking the insert. This fragment was subcloned into pBluescript SK⁺ (Stratagene). Nested unidirectional deletions were prepared²⁵ and sequenced²⁶ using modified T7 DNA polymerase (Sequenase, version 2; United States Biochemical Corp.)²⁷ and buffer-gradient gels²⁸. The sequence was determined on both strands.

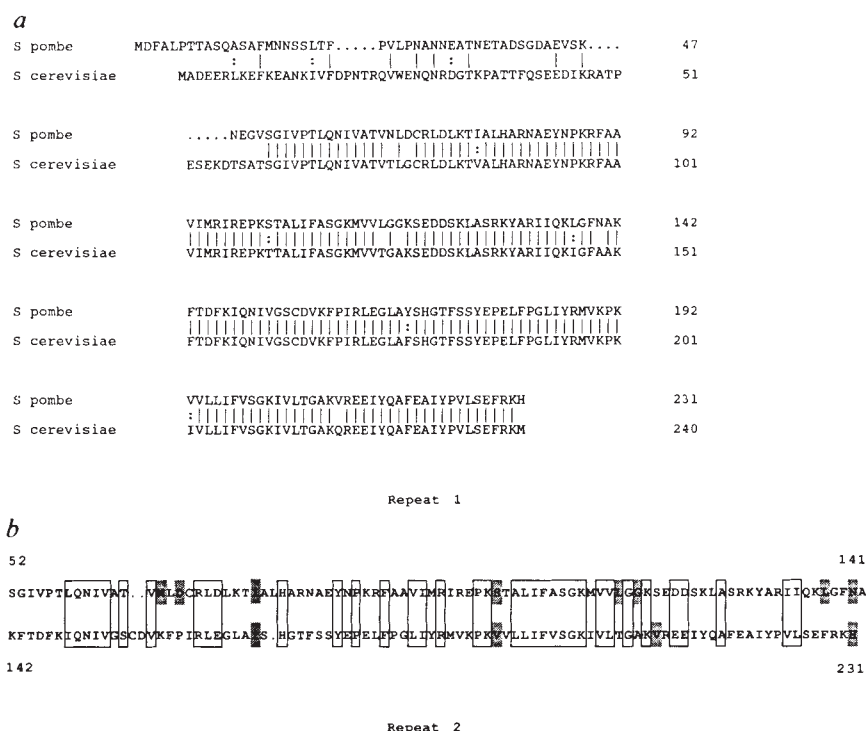
pombe and *S. cerevisiae* is 83%, 82%, 93% and 91% for H2A, H2B, H3 and H4, respectively¹⁸. Other proteins, such as alcohol dehydrogenase and triose phosphate isomerase, exhibit far greater divergence between the two organisms (50% and 59% homology)^{19,20}.

The remarkable evolutionary conservation of amino-acid sequence underscores the importance of the carboxy-terminal 178 amino acids of the TFIID protein. We note that this conserved region corresponds to that segment of the *S. cerevisiae*

protein in which a directly repeated motif was previously discerned^{15,21,22}. Figure 4b aligns the repeats of *Schiz. pombe* TFIID. Without exception, residues conserved between the *Schiz. pombe* repeats (36/89) are invariant between the two species. Thus, the structure conferred by these direct repeats is undoubtedly important for TFIID function.

Eight independent isolates of TFIID (class I) and two isolates of *Lys2* (class II) were obtained in a single screen of the library. Although we have not identified any other cDNAs from *Schiz.*

FIG. 4 Analysis of *Schiz. pombe* TFIID predicted amino-acid sequence. *a*, Alignment of the predicted amino-acid sequence of *S. pombe* TFIID with that of *S. cerevisiae* TFIID reveals weak homology at the amino terminus, but near-identity in the final 178 amino acids. Identity between residues is shown as a vertical line, conservative changes are identified by two dots, according to the groupings (K,R,H), (E,D), (N,Q), (G,P), (A,I,L,V), (F,W,Y), (M,C) and (S,T) (one-letter code). Optimal alignment of amino-terminal sequences required the introduction of two gaps in the *Schiz. pombe* sequence. *b*, The region of *Schiz. pombe* TFIID with strongest homology to *S. cerevisiae* TFIID contains a directly repeated sequence. Residues that are identical or conserved (as defined above) between the two aligned, contiguous regions are boxed. Differences from the *S. cerevisiae* sequence are shaded, and are found without exception outside the repeated residues.



pombe that will complement the *spt15-21* mutation in FW1259, our approach of functional cloning using this comprehensive library will enable us to search exhaustively for any genes encoding minor forms of TFIID-like proteins that may be expressed at low levels. □

Received 5 April; accepted 15 May 1990.

1. Forsburg, S. & Guarente, L. *A. Rev. Cell. Biol.* **5**, 153-180 (1989).
2. Struhl, K. *Cell* **49**, 295-297 (1987).
3. Johnson, P. & McKnight, S. A. *Rev. Biochem.* **58**, 799-839 (1989).
4. Mitchell, P. J. & Tjian, R. *Science* **228**, 371-378 (1989).
5. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. *Cell* **56**, 549-561 (1989).
6. Hawley, D. & Roeder, R. *J. Biol. Chem.* **262**, 3452-3461 (1987).
7. Van Dyke, M. W., Roeder, R. G. & Sawadogo, M. *Science* **241**, 1335-1338 (1988).
8. Buratowski, S., Hahn, S., Sharp, P. & Guarente, L. *Nature* **334**, 37-42 (1988).
9. Cavallini, B. *et al. Nature* **334**, 77-80 (1988).
10. Horikoshi, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 4843-4847 (1989).
11. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. *Cell* **58**, 1173-1181 (1989).
12. Eisenmann, D. M., Dollard, C. & Winston, F. *Cell* **58**, 1183-1191 (1989).
13. Horikoshi, M. *et al. Nature* **341**, 299-303 (1989).

14. Schmidt, M. C., Kao, C. C., Pei, R. & Berk, A. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7785-7789 (1989).
15. Cavallini, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 9803-9807 (1989).
16. Ammerer, G. *Meth. Enzym.* **101**, 192-201 (1983).
17. Winston, F., Chaleff, D., Valent, B. & Fink, G. *Genetics* **107**, 179-197 (1984).
18. Matsumoto, D. & Yanagida, M. *EMBO J.* **4**, 3531-3538 (1985).
19. Russell, P. & Hall, B. D. *J. Biol. Chem.* **258**, 143-149 (1983).
20. Russell, P. R. *Gene* **40**, 125-130 (1985).
21. Hoeijmakers, J. J. *Nature* **343**, 417-418 (1990).
22. Nagai, K. *Nature* **343**, 418 (1990).
23. Sherman, F., Fink, G. R. & Hicks, J. B. *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, New York, 1986).
24. Rothstein, R. J. *Meth. Enzym.* **101**, 202-211 (1983).
25. Henikoff, S. *Gene* **28**, 351-359 (1984).
26. Del Sal, G., Manfioletti, G. & Schneider, D. *Biotechniques* **7**, 514-519 (1989).
27. Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4767-4771 (1987).
28. Sheen, J.-Y. & Seed, B. *Biotechniques* **6**, 942-944 (1988).

ACKNOWLEDGEMENTS. This work represents equivalent efforts by J.D.F. and D.M.B., who should be considered equally as first author. We thank Biorad for loan of an electroporator, Drs David Eide and Gerry Fink for strains, and Drs Eide and Olivia Bermingham-McDonogh for comments on the manuscript. D.M.B. is indebted to Dr A. Lee Osterman (Hospital of the University of Pennsylvania) for his expertise. This work was supported by grants from the NIH (L.G. and F.W.) and postdoctoral fellowships from the American Cancer Society (J.D.F. and D.M.B.). The nucleotide sequence reported here has been deposited in the EMBL, GenBank and DDBJ databases under accession number X53383.

Atomic-scale imaging of DNA using scanning tunnelling microscopy

Robert J. Driscoll, Michael G. Youngquist
& John D. Baldeschwieler*

Division of Chemistry and Chemical Engineering,
Arthur Amos Noyes Laboratory of Chemical Physics, Mail Code 127-72,
California Institute of Technology, Pasadena, California 91125, USA

THE scanning tunnelling microscope (STM) has been used to visualize DNA¹ under water², under oil³ and in air⁴⁻⁶. Images of single-stranded DNA have shown that submolecular resolution is possible⁷. Here we describe atomic-resolution imaging of duplex DNA. Topographic STM images of uncoated duplex DNA on a graphite substrate obtained in ultra-high vacuum are presented that show double-helical structure, base pairs, and atomic-scale substructure. Experimental STM profiles show excellent correlation with atomic contours of the van der Waals surface of A-form DNA derived from X-ray crystallography. A comparison of variations in the barrier to quantum mechanical tunnelling (barrier-height) with atomic-scale topography shows correlation over the phosphate-sugar backbone but anticorrelation over the base pairs. This relationship may be due to the different chemical characteristics of parts of the molecule. Further investigation of this phenomenon should lead to a better understanding of the physics of imaging adsorbates with the STM and may prove useful in sequencing DNA. The improved resolution compared with previously published STM images of DNA may be attributable to ultra-high vacuum, high data-pixel density, slow scan rate, a fortuitously clean and sharp tip and/or a relatively dilute and extremely clean sample solution. This work demonstrates the potential of the STM for characterization of large biomolecular structures, but additional development will be required to make such high resolution imaging of DNA and other large molecules routine.

We have used a Caltech-constructed ultra-high vacuum (UHV) STM system similar to the IBM 'pocket-size' STM⁸. On gallium arsenide our vertical resolution has been as high as ~1 pm (unpublished data). The system incorporates *in vacuo* tip and sample transfer and has a base pressure of ~3 × 10⁻¹¹ torr.

Although UHV provides a clean experimental environment it may cause some DNA denaturation; intact DNA is probably

TABLE 1 Comparison of DNA dimensions derived from the STM and X-ray crystallography

	STM	X-ray
Helix pitch	29 Å	28.5 Å
Minor groove width	10 Å	11.0 Å
Major groove width	3 Å	2.7 Å
Molecular width	23 Å	23 Å
Phosphate backbone width	10 Å	11.6 Å
Axial nucleotide rise	2.6 Å	2.59 Å
Base-pair angle	+18°	+19°
Helix symmetry	11	11.0

Comparison of the dimensions of general features in the STM image and dimensions derived from X-ray crystallography for 'random-sequence' A-DNA. The STM dimensions are average approximate values based on Fig. 1a. STM differs from X-ray crystallography in its ability to study local dimension and orientation variations on an individual molecule. Crystallography data reflect the average dimensions of many molecules in a crystal and therefore cannot be used to study local structural variations.

in an A-type conformation⁹. The conventional A form, determined by X-ray crystallography, is characterized by a width of ~23 Å, helical symmetry of 11 base pairs per turn, a +19° base-pair tilt to the helix axis, a 28.5 Å pitch and an axial nucleotide rise of 2.59 Å (ref. 9). Subtracting the van der Waals width (11.6 Å) of the backbone phosphate groups reveals that the major groove is narrow and deep (2.7 Å and 13.5 Å, respectively) and the minor groove is wide and shallow (11.0 Å and 2.8 Å).

Figure 1a shows an image of DNA from our UHV STM. The double helix and the major-minor groove alternation are apparent, as are parallel features spanning the minor grooves at a +18° ± 3° angle to the helix axis. Their orientation and spacing are consistent with base pairs bridging the thick phosphate backbones (~10 Å wide; instrument calibrated using the graphite atomic lattice). Roughly four turns of the helix are shown; the helix symmetry estimated from the bases visible in the minor grooves is ~11 base pairs per turn, consistent with A-DNA. The bottom portion of Fig. 1a shows a wide (~10 Å) minor groove with considerable substructure leading into a narrow (~3 Å), less-resolved major groove, consistent with the structure of A-DNA. The major groove is too narrow for the STM tip to enter. During a scan line in the second minor groove, a tip instability occurred. After this event the DNA molecule looks slightly different; the major groove appears as a notch on the right side of the molecule in the transition between one minor groove to the next. This may be explained by the tip having moved the upper half of the DNA slightly. But this same change in orientation is evident in previous images and probably

* To whom correspondence should be addressed.

FIG. 1 *a*, Unsmoothed, unfiltered plane-subtracted STM image of DNA $\sim 80 \times 120 \text{ \AA}$ (400×250 data pixels). Yellow, topographic maxima. The parallel bands at $18^\circ \pm 3^\circ$ to the helix axis are base pairs in the minor grooves. Several of the base pairs crossing one of the minor grooves are indicated by red arrows. During a brief instability (yellow arrow) the probe tip may have tilted the molecule. *b*, Model of the van der Waals surface of A-DNA derived from X-ray crystallographic data, scaled to *a*. The sugar-phosphate backbone (red) and the base pairs (yellow). Hydrogen atoms omitted for clarity. Two distinct orientations differing by only a few degrees are illustrated with the transition (arrow) chosen to coincide with the tip instability seen in *a*. Model constructed using BioDesign Biograf software.

METHODS. The sample was a ~ 550 base pair (bp) *Avall* fragment of mouse B-cell V-region DNA (from pBV13L; Jerry Siu, Caltech). A $2\text{-}\mu\text{l}$ drop of 10 mM aqueous ammonium acetate solution containing 10 ng DNA per μl was

deposited on freshly cleaved highly oriented pyrolytic graphite (HOPG; Union Carbide, zya grade) in air and dried in vacuum. No visible residue was present. The scan rate was $\sim 100 \text{ \AA s}^{-1}$ with sample bias of +100 mV, feedback-maintained tunnelling current of 1 nA, and system pressure of $< 10^{-10}$ torr. The probe tip was electrochemically etched tungsten (6 V a.c., 2 M KOH).

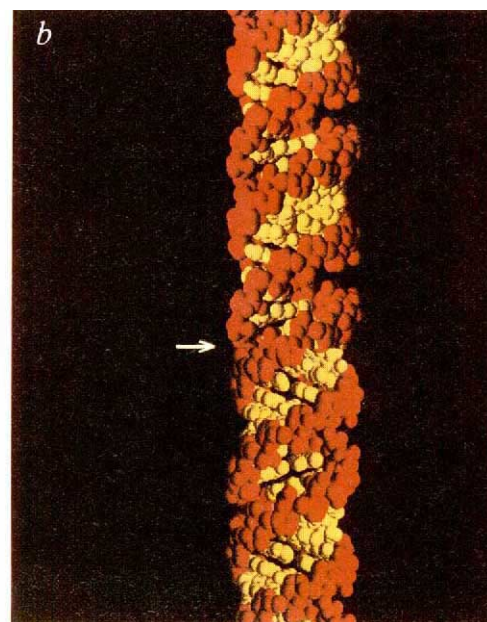
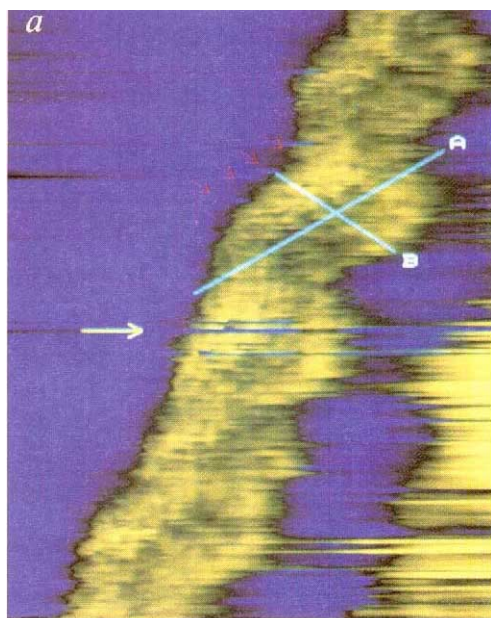


FIG. 2 Comparison of the bottom portion of Fig. 1*a* with a corresponding section of the van der Waals' model of A-DNA. *a*, STM image is $\sim 35 \times 55 \text{ \AA}$ and shows nearly two helical turns. The colour coding is identical to that in Fig. 1*a*. The plane-subtracted data are unfiltered and unsmoothed, but contrast has been enhanced by histogram equalization. The *y* axis is skewed by several degrees to facilitate direct comparison. *b*, The van der Waals' model shows both backbone and bases in yellow. Hydrogen atoms are omitted.

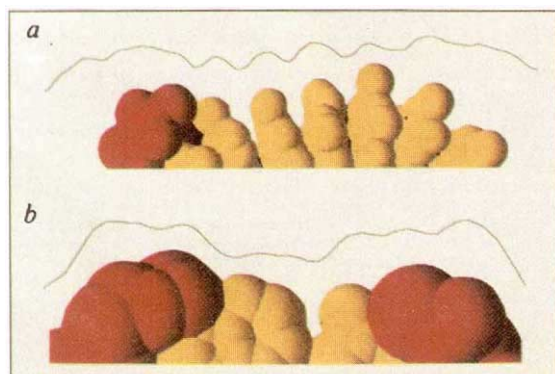
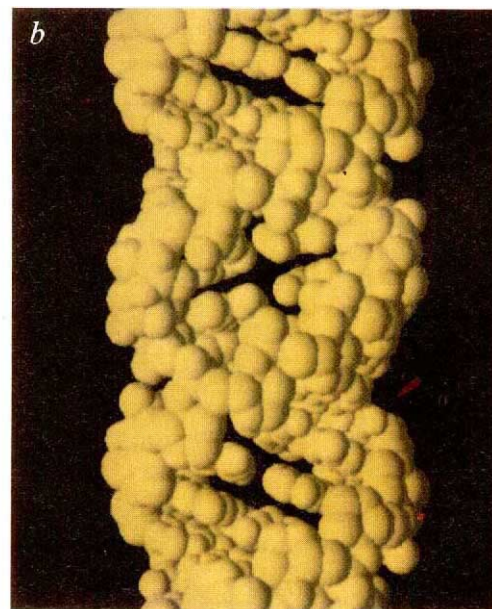
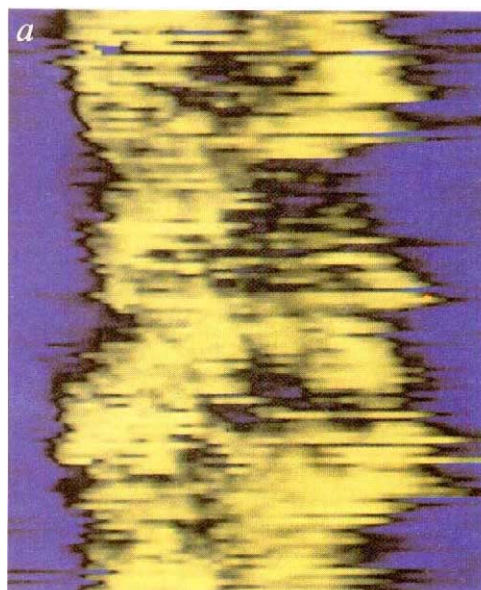


FIG. 3 Interpolated experimental STM tip trajectories following lines marked A and B on Fig. 1*a* compared with the corresponding atomic contours of an A-DNA van der Waals surface model. In each case the experimental cross-section is placed above the corresponding region of the model. *a*, Cut across the base-pair planes approximating the minor groove axis (line A in Fig. 1*a*). Starting at the left, the line cuts through the phosphate-sugar backbone on the leading edge and shows the base-pair periodicity across the groove. *b*, Similar comparison (line B in Fig. 1*a*) across the minor groove through the two high backbones and an intermediate base pair. Height and length are on the same scale. The data were smoothed using a binomially weighted sliding window average, corresponding to a two-dimensional gaussian of $\sim 0.60 \text{ \AA}$ FWHM (small compared with the typical van der Waals diameter of $3 \text{ \AA}$). Although hydrogen atoms are not shown, STM may be influenced by their presence.

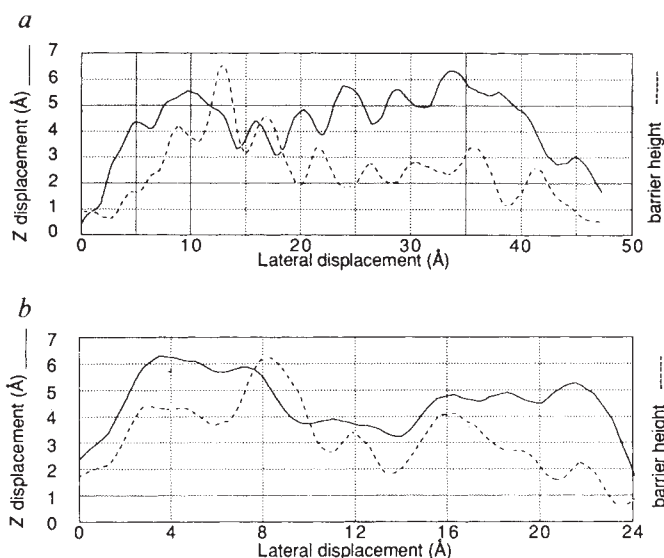


FIG. 4 Interpolated cross-sections from the topographical and barrier-height data of Fig. 1a. Solid lines, tip trajectories as shown in Fig. 3; dashed lines, corresponding data from a simultaneously acquired gap-modulated barrier-height image. Barrier-height axes are in arbitrary units. The effective local barrier height was measured using the standard method of sinusoidally modulating the tip-sample separation ($0.67 \text{ \AA}$ p-p at 1 kHz) and detecting $d(\ln i)/ds$, the derivative of the natural logarithm of the tunnelling current (i) with respect to the tip-sample separation, with a lock-in amplifier.

represents a kink site. A tilting of a few degrees can result in this orientation, as is demonstrated with the model of the A-DNA van der Waals surface shown in Fig. 1b.

The average width of the molecule in the STM image is $\sim 23 \text{ \AA}$, and the average 'apparent height'—the distance the tip moved up over the DNA—is $\sim 12 \text{ \AA}$. The image acquired scanning in the opposite direction in x shows similar structure, but the molecule appears slightly narrower ($\sim 21 \text{ \AA}$). This may indicate elastic interaction between the tip and the molecule. The DNA molecule might be expected to appear slightly broadened in STM images because of the size and curvature of the tip on the scale of DNA; the edge of the tip will probably sense the $20\text{-\AA}$ -high DNA molecule before the lowest point on the tip. A shortened 'height' of the molecule is also not surprising; the current is dependent on both the tip-sample separation and the effective tunnelling barrier. In the topographic imaging mode used here, the STM maintains a constant electron tunnelling current and therefore a constant tip-sample separation if the effective local tunnelling barrier-height is constant. In response to an increased barrier-height over the molecule, however, the feedback system must reduce tip-sample separation to maintain constant current. The dimensions of the DNA in the image are in general agreement with those of conventional A-DNA, such that the average pitch is $\sim 29 \text{ \AA}$ and the axial nucleotide rise is $\sim 2.6 \text{ \AA}$. These dimensions and those derived from X-ray crystallography are compared in Table 1. Any differences may be explained by the variability in the structure of A-family polynucleotides, substrate-tip thermal drift, or piezo creep. In addition, it is not expected that DNA dried in vacuum on a substrate would have a structure fully consistent with X-ray crystallographic data.

A direct comparison of the bottom portion of Fig. 1a and an A-DNA model representing a corresponding section is provided in Fig. 2. Starting at the bottom of Fig. 2a, the alternating minor-major-minor groove sequence is readily apparent. The varying thicknesses of the backbones on the surface of the DNA molecule are visible, as are two well-resolved base pairs in the bottom minor groove. There is good agreement in the general

features between the STM image and the model, and the image suggests atomic features. To assess the presence of features on the atomic scale, interpolated cross-sections of the topography were chosen to span the minor groove showing the best resolution (see lines A and B in Fig. 1a) and are shown in Fig. 3. The almost atom-for-atom agreement of the experimental data cross-sections with the contours of the models in both Fig. 3a and b indicates atomic resolution in this region of the image. The ability to resolve atomic features on a $>20 \text{ \AA}$ -diameter biomolecule deserves further investigation. Models proposed to explain the STM imaging of DNA did not anticipate STM response to the atomic topography of the molecule^{5,10,11}. This result requires a separation-dependent current that is sensitive to the topographic surface of the molecule on the atomic scale and should be considered in the development of more-complete image-contrast theories.

Interpolated cross-sections of the topographical and barrier-height data from the image in Fig. 1a are compared in Fig. 4. The barrier-height image is not shown here but shows features similar to the topographical image. Although the barrier-height data are shown in arbitrary units, we typically measure effective local barrier-heights over DNA ranging $0.2\text{--}1.5 \text{ eV}$. The topographical and barrier-height features show fair correlation of structure over the backbone atoms. But the barrier-height reveals considerable anti-correlation (complementarity) in the minor groove where atoms of the base pairs are resolved; the barrier-height shows a peak where there is a valley in the topographical cross-section and vice versa. This complementarity is quite pronounced over the base-pair atoms in Fig. 4a. The topography in Fig. 4b shows two high backbones with a concave minor groove of depth $2\text{--}3 \text{ \AA}$ and width $\sim 10 \text{ \AA}$, again consistent with the van der Waals surface model. The agreement of features over the backbone and complementarity over the bases is also evident in this cross-section. Complementarity may be due to the stronger influence of chemical identity in barrier-height imaging. Current imaging tunnelling spectroscopy (CITS) may reveal additional chemical information¹², but our preliminary measurements show little difference in current-voltage character between the bare substrate and the adsorbed molecule.

These relationships and the demonstrated ability to image atomic-scale features of a biomolecule raise interesting questions about the mechanism of STM imaging of a biological structure, such as the conditions necessary for the STM to respond to the atomic topography of a large molecule, the factors causing barrier-height and topography to be in phase over some parts of the molecule and out of phase over others and the separation-dependent conduction mechanism giving rise to the contrast in STM images of DNA. Possible mechanisms will be discussed elsewhere. Further investigations of the sensitivity of the barrier-height and CITS measurements to the identity of atomic or functional groups may lead to new techniques for identifying such groups. With substantial technological advances, sequencing of DNA by STM may be possible. $\square$

Received 6 February; accepted 23 April 1990.

1. Binnig, G. & Rohrer, H. in *Trends in Physics* (eds Janta, J. & Pantoflick, J.) 38–46 (European Physical Society, The Hague, 1984).
2. Lindsay, S. M., Thundat, T., Nagahara, L., Knipping, U. & Rill, R. L. *Science* **244**, 1063–1064 (1989).
3. Arscott, P. G., Lee, G., Bloomfield, V. A. & Evans, D. F. *Nature* **339**, 484–486 (1989).
4. Cricenti, A. et al. *Science* **245**, 1226–1227 (1989).
5. Amrein, M., Dürr, R., Stasiak, A., Gross, H. & Travaglini, G. *Science* **243**, 1708–1711 (1989).
6. Beebe, T. P. Jr et al. *Science* **243**, 370–372 (1989).
7. Dunlap, D. D. & Bustamante, C. *Nature* **342**, 204–206 (1989).
8. Gerber, Ch., Binnig, G., Fuchs, H., Marti, O. & Rohrer, H. *Rev. Sci. Instrum.* **57**, 221–224 (1986).
9. Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).
10. Keller, D., Bustamante, C. & Keller, R. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5356–5360 (1989).
11. Lindsay, S. M., Sankey, O. F., Li, Y., Herbst, C. & Rupprecht, A. *J. Phys. Chem.* (in the press).
12. Tromp, R. M., Hamers, R. J. & Demuth, J. E. *Science* **234**, 304–309 (1986).

ACKNOWLEDGEMENTS. R.J.D. and M.G.Y. are both to be considered as first author. We thank C. Spence and J. Meier of Dr L. Hood's group for persuading us to undertake this project and for preparing the DNA solution. We also thank J. Hurley and M. Sud of BioDesign Inc. for providing the molecular modelling software, and Dr W. A. Goddard III for discussions of this work. R.J.D. is an NIH trainee and M.G.Y. is a Department of Education fellow. This work has been supported by the NIH, ONR and the Shell Companies Foundation.

Light-based detection of biomolecules

I. Durrant

Horseradish peroxidase-labelled nucleic acid or antibody probes detected with enhanced chemiluminescence offer rapid, convenient and sensitive systems for membrane-based detection procedures in molecular biology.

THE detection of DNA, RNA and proteins on membranes has traditionally involved the incorporation of radioactive labels. Such systems still represent the most sensitive and robust techniques available. However, there is an increasing demand for non-radioactive systems that offer similar sensitivity but avoid the licensing procedures that are associated with the handling and disposal of radioactive materials.

Ideally, such systems should offer high sensitivity, reproducible results, rapid labelling and detection procedures, an ability to reprobe target sequences and should yield a permanent 'hard copy' of the result. Until recently, commercially available non-radioactive systems¹ have suffered from several inherent disadvantages. These include poor detection sensitivity, inconvenient labelling and washing steps and the use of colorimetric substrates that are incompatible with multiple reprobing experiments.

Two technologies have now been combined enabling these criteria to be more closely met. These are (1) the direct labelling of probes with enzymes (for example, horseradish peroxidase (HRP))² and (2) the concept of luminol-based enhanced chemiluminescence (ECL) catalysed by HRP. This enzyme-based detection system can be used with long nucleotide sequences (DNA or RNA) or oligonucleotides. For protein detection, antibodies can be conjugated to HRP and detected using ECL.

Labelling strategies

The three potential probe formats require three separate labelling strategies. The conjugation of HRP to antibodies is well documented³ and will not be discussed further.

Direct labelling of long nucleotide sequences with HRP is achieved by the method originally developed by Renz⁴ at the European Molecular Biology Laboratory. The labelling reagent is a positive-charge modified, polymerized HRP complex consisting of 2–3 HRP molecules covalently cross-linked, via *para*-benzoquinone, to polyethyleneimine (PEI). This complex interacts electrostatically with negatively-charged, single-stranded nucleic acid at a frequency of approximately one complex every 30 bases. The interaction is stabilized by addition of glutaraldehyde, which introduces covalent cross-links between amino groups on PEI and the nucleotides. The

entire procedure takes 20 minutes and labelled probes, stored at -20°C in 50 per cent glycerol, are stable for at least 9 months. The chemical reactions and interactions are reproducible provided that the salt concentration in the nucleic acid solution is less than 10 mM and double-stranded probes are rendered completely single-stranded before labelling, for example, by use of a vigorously boiling water bath.

The direct labelling of oligonucleotides uses a linker compound that incorporates a reactive thiol group at the 5' end of a sequence using an automated oligonucleotide synthesizer. Derivatized HRP reacts readily with the thiol group on the oligonucleotide. The labelling reaction is simplified by supplying lyophilized, derivatized HRP in the reaction tube. The thiol-oligonucleotide is reacted with a five-fold molar excess of derivatized HRP for 1 hour at room temperature. Because of the high efficiency of this reaction the conjugate requires no further purification and the free HRP does not appear to interfere with subsequent hybridizations. Labelled probes, stored as discussed for long probes, have a similar stability of at least 9 months.

Hybridization step

The overriding factor when handling directly-labelled probes is to maintain enzyme activity. Consequently, the common feature of the hybridization conditions is a maximum temperature of 42°C .

For long probes, hybridizations are performed at a probe concentration of $10\text{--}20\text{ ng ml}^{-1}$, in a solution using 6 M urea as a denaturant; 6 M urea buffers act in a similar manner to those containing 50 per cent formamide. In addition, the buffer contains a rate enhancer and an enzyme stabilizing agent. In most situations, a final salt concentration of 0.5 M is optimal, yielding the maximum rate of hybridization⁵. Stringency can be controlled by varying the salt concentration in the stringency wash buffer. Such buffers also contain 6 M urea and should not be used above 42°C . For high-sensitivity applications, over-night hybridizations are required. However, for high target systems — colony and plaque lifts, detection of amplified sequences — hybridization times of 2–4 hours are possible.

Enzyme-labelled oligonucleotide probes are hybridized for 1–2 hours at a concentration of 20 ng ml^{-1} in a buffer

based on salt, sodium citrate and detergent. Stringency is controlled in the post-hybridization washes and can be achieved by the inclusion of urea or formamide, by variation in the salt concentration, by an increase in wash temperature or by a combination of these factors. The addition of formamide or the increase in temperature above 42°C can cause loss of HRP activity, but these problems will not significantly affect the final result if relatively short incubations are used. Stringency can be controlled to such a degree that probes with one mismatch in 16 bases can be distinguished from perfectly-matched probes⁶.

The inclusion of a proteinaceous blocking agent in the relevant hybridization buffer overcomes the tendency of enzyme-labelled probes to bind non-specifically to nylon membranes.

Detection system

The oxidative degradation of a variety of compounds is known to yield chemiluminescence. The most clearly understood system is the HRP/hydrogen peroxide catalysed oxidation of luminol in alkaline conditions⁷ (see Fig. 1). The luminol radicals formed at reactions 2 and 3 inter-

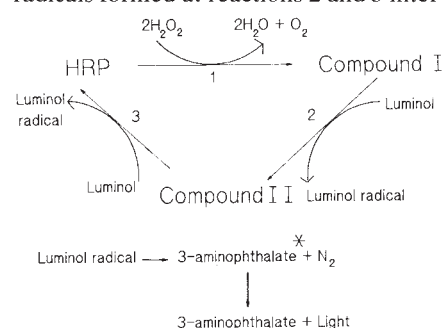


FIG. 1 Proposed reaction scheme for the HRP-catalysed oxidation of luminol leading to the emission of light.

act in a series of steps, involving oxygen from reaction 1, to yield an excited form of 3-aminophthalate. As this compound relaxes to the ground state, light is emitted. Light output is restricted by the rate-limiting step in this cycle — reaction 3. However, Whitehead *et al.*⁸ demonstrated that certain compounds were able to enhance the reaction in terms of light output by 10–100 fold. Extensive studies of the mechanism of this phenomenon have led to the use of *para*-substituted phenolic compounds as enhancers⁹. These are capable of increasing light output by greater than 1000-fold and reducing the background by inhibiting chemical oxida-

tion of luminol. The enhancer is thought to act by replacing luminol as a substrate for reaction 3, thus removing the rate-limiting step. The enhancer radicals formed react with luminol so that the light-producing pathway is the same in the enhanced and unenhanced reactions.

Consequently, a peracid salt is used as a source of peroxide and is stored separately from the luminol/enhancer mixture. In this form, the substrates are stable for at least 6 months and the two solutions are mixed immediately before use. Typical exposure times are 1–60 minutes (see Fig. 2) depending on the level of the target (either nucleic acid or antigen). Light emission on membrane actually peaks 5–

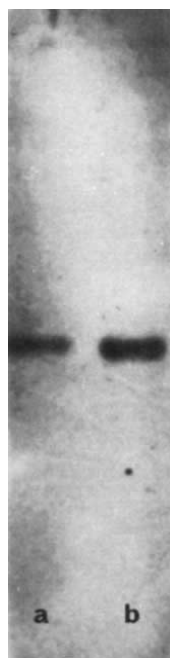


FIG. 2 Single-copy gene detection in an *EcoRI* digest of human genomic DNA blotted on to Hybond N^+ (nylon membrane). Gel loadings of 2 and 5 μg (tracks a and b). Overnight hybridization with a 3-kb, HRP-labelled probe for *mos* proto-oncogene (probe stored for 6 months before use). Probe concentration 20 ng ml^{-1} , film exposed for 30 minutes.

10 minutes after addition of substrate and lasts for 2–3 hours. The eventual total loss of signal from the probe eliminates the need physically to strip this probe from the target before subsequent reprobings. Indeed, the same probe can be used repeatedly as the first probe can exchange with the second. Up to ten reprobings are possible when using nylon membranes without significant loss in sensitivity.

The light is emitted with a maximum wavelength of 428 nm making it ideal for capture on X-ray film which is designed to detect the blue light emitted by standard intensifying screens. The best results are obtained on grey-tinted, rather than blue-tinted, films. Alternatively, a charge-coupled device (CCD) camera¹⁰ may be used allowing accurate quantification of the number of photons of light emitted over five orders of magnitude.

Sensitivity of detection

When used with long probes, the ECL detection system has a sensitivity of at least 1×10^{-18} mols of target (depending on probe length), which is the target level of a single-copy gene in 2 μg of a human

Labelling technology

A random-primed DNA labelling kit that uses exonuclease-free Klenow, a DNA typing kit, and immunogold-conjugated F(ab')_2 fragments are among this week's products that focus on labelling techniques.

British Bio-technology has developed a new range of **fluorokine reagents** in kit form, which are designed for labelling and studying cytokine receptors on cell surfaces by flow cytometry (*Reader Service No. 101*). At present, kits to $\text{IL-1}\alpha$, IL-2 , IL-3 , IL-4 , IL-6 , GM-CSF and G-CSF receptors are available. The kits contain biotinylated or fluorescently-labelled cytokines, detection reagents and wash buffers that are optimized to lower non-specific binding and to ensure the sensitivity necessary to study the low number of cytokine receptors that are usually expressed on cell surfaces. The labelled cytokine binds to its receptor with high specificity, such that fluorescent intensity is directly proportional to the density of cell surface receptors, says BBT. In addition, the kits provide a means of determining the percentage of cytokine receptor-bearing cells in mixed-cell populations.

Delfia labelling reagents and kits from Pharmacia provide a practical alternative to the use of radioactive labels for the **labelling of proteins, cell markers and DNA probes** (*Reader Service No. 102*). The kits are based on the unique fluorescent properties of lanthanide metal chelates, which, as fluorophores, have a large Stokes shift and a long decay time. Time-resolved fluorometric measurements can, therefore, be made at a wavelength and time where excitation light and non-specific background fluorescence are



Delfia reagents and kits for non-radioactive multiple labelling.

minimal, says Pharmacia. Europium, samarium and terbium all have discernable emission peaks and so applications requiring dual or triple labelling are possible. The labelling procedure itself is mild. As it does not involve the use of denaturing compounds, there is minimal influence on biological activity or immunoreactivity. Delfia technology offers labels that are stable for up to 12 months.

Oxford GlycoSystems has extended its range of **labelled oligosaccharides** (*Reader Service No. 103*). Included in the range are tritium-labelled alditol derivatives, and biotin-, fluorescein- and dansyl-labelled oligosaccharides. The tritium-labelled alditols are made by reduction of the parent oligosaccharide with sodium borotritide under optimal conditions to maximize specific activity and eliminate base-induced degradation, says the company. The non-radioactive labelled oligosaccharides are substituted at the reducing

genomic DNA digest¹¹. The oligonucleotide system has a sensitivity limit of $10\text{--}20 \times 10^{-18}$ mols. Both systems are suitable for detecting target sequences in various applications including Southern and northern blots, colony and plaque screening, fingerprint analysis and the detection of PCR-amplified sequences¹². The detection of antigen on western blots using HRP-conjugated antibodies and ECL detection offers both rapid results and a sensitivity that is at least ten-fold greater than colour-producing substrates.

In conclusion, these non-radioactive systems offer high sensitivity, rapid and reliable labelling and detection procedures, permanent 'hard-copy' results on X-ray film and simple protocols for reprobings membranes. □

Ian Durrant is a Senior Development Scientist in the DNA Labelling and Detection Programme, Amersham International plc, White

Lion Road, Amersham, Buckinghamshire, England HP7 9LL, UK. For more information, fill in reader service number 100.

- Pollard-Knight, D. *Technique* **2**, 1–20 (1990).
- ECL Technical Manual available on request from Amersham International plc, Amersham, UK (1990).
- Hashida, S., Imagawa, M., Inoue, S., Ruan, K. & Ishikawa, E. *J. appl. Biochem.* **6**, 56–63 (1984).
- Ren, M. & Kurz, C. *Nucleic Acids Res.* **12**, 3435–3444 (1984).
- Pollard-Knight, D. et al. *Analyt. Biochem.* **185**, 84–89 (1990).
- Evans, M.R. et al. *Proc. 6th int. Congr. on Rapid Methods and Automation in Microbiol. & Immunol.* (in the press) (1990).
- Thorpe, G.H.G. & Kricka, L.J. in *Bioluminescence and Chemiluminescence: New Perspectives* (eds Scholmerich, J. et al.) (Wiley, Chichester, 1987).
- Whitehead, T.P., Thorpe, G.H.G., Carter, T.J.N., Groucutt, C. & Kricka, L.J. *Nature* **305**, 158–159 (1983).
- Hodgson, M. & Jones, P. *J. Biolum. Chemilum.* **3**, 21–25 (1989).
- Boniszewski, Z.A.M., Comley, J.S., Hughes, B. & Read, C.A. *Electrophoresis* (in the press) (1990).
- Durrant, I. et al. *Biotechniques* **8**, 564–570 (1990).
- Sorg, R., Enczmann, J., Sorg, U., Kogler, G., Schneider, E.M. & Wernet, P. *Life Science* **2** available on request from Amersham International plc, Amersham, UK (in the press) (1990).

terminus with an amino acid spacer arm carrying the label. In this form, the oligosaccharide remains intact and retains all the features of N-linked oligosaccharides on the glycoprotein.

Over 500 new compounds are offered in Merck Sharp & Dohme's **1990 catalogue of stable isotopes** (Reader Service No. 104). The catalogue includes expanded sections on deuterium, carbon-13 and nitrogen-15 labelled compounds, with new sections covering labelled t-BOC amino acids and drug standards. New algal-derived labelled products include the lower levels of carbon-13 labelling for two-dimensional experiments using nuclear magnetic resonance. Additional spin traps and spin labels for ESR spectroscopy are also featured in the catalogue.

Probing matters

Boehringer is featuring two non-radioactive **DNA and RNA labelling kits** and a nucleic acid detection kit, which are based on digoxigenin technology (Reader Service No. 105). Labelled DNA probes are generated with Klenow polymerase by random-primed incorporation of digoxigenin-labelled deoxyuridine triphosphate (dUTP). The dUTP is linked via a spacer arm to digoxigenin (DIG-dUTP). Labelled RNA probes are synthesized by *in vitro* transcription of DNA, cloned downstream of SP6, T7 or T3 promoters, with SP6, T7 or T3 RNA polymerases using digoxigenin-labelled uridine triphosphate (DIG-UTP) as a substrate. Both DIG-labelled DNA or RNA probes are detected, after hybridization to target nucleic acids, by enzyme-linked immunoassay using an antibody conjugate (anti-digoxigenin alkaline phosphatase conjugate). Detection is by the formation of an insoluble blue precipitate. Boehringer's digoxigenin DNA and RNA labelling and detection system is suitable for DNA and RNA transfers, colony or plaque screening and *in situ* hybridizations.

United States Biochemical has modified its **random-primed DNA labelling kit** to include exonuclease-free Klenow (Reader Service No. 106). This, says USB, offers several advantages over the use of regular Klenow: probes made with exonuclease-free Klenow under standard reaction conditions are 2–3-fold longer; total incorporation of labelled nucleotide is improved by a factor of 1.5-fold; and biotinylated nucleoside triphosphates are incorporated to a greater extent in the labelling reaction over a longer period of time. The \$100 (US) random-primed DNA labelling kit provides all the reagents, except for labelled deoxynucleoside triphosphate (^{32}P , ^{35}S , ^3H , biotin-labelled), necessary for 50 labelling reactions. Probes labelled in this way are suitable for all types of hybridization

techniques, including Southern and northern blots and *in situ* hybridizations.

GenePrint is a non-radioactive **DNA typing kit** from Promega, which offers a convenient system for the chemiluminescent detection of digoxigenin-labelled DNA probes (Reader Service No. 107). Although designed primarily for paternity testing and cell-line authentication, GenePrint can, however, be used in any application where sensitive non-radioactive detection is required. Digoxigenin-labelled VNTR probes can be purchased with the kit or, alternatively, the kit can be used with any digoxigenin-labelled probe. The \$543 (US) GenePrint kit includes chemiluminescent substrate, alkaline phosphatase antibody conjugate, analytical



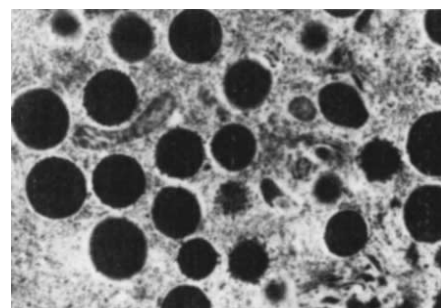
DNA typing with GenePrint.

markers, allelic controls, restriction enzyme, nylon membrane, protocol and technical manual. Each kit contains enough reagents to DNA type 100 samples, says Promega.

A range of enhanced chemiluminescence (ECL) **non-isotopic light detection systems** are now available from Amersham (Reader Service No. 108). With the ECL systems, horseradish peroxidase is used to label nucleic acids, oligonucleotides or proteins. The blue light produced in the detection reaction (oxidation of luminol) is detected on X-ray film. The ECL gene detection system for nucleic acid probes of greater than 50 bp is, says Amersham, fast and easy to use: labelling, hybridization and detection on nylon or nitrocellulose membranes takes less than seven hours. The ECL oligonucleotide labelling and detection system takes less than five hours and target levels down to 20×10^{-18} mols can be detected. The ECL western blotting detection system can detect protein targets of less than 1 pg in 15 seconds to 60 minutes. The kit, which contains sufficient reagents to process 4,000 cm² of blots, is 10 times more sensitive than colorimetric systems, says Amersham.

Ideas for immunology

Caltag Laboratories has launched a new line of **immunogold conjugates** for electron microscopy, light microscopy and



Immunogold conjugates — just the job for EM, light microscopy and immunoblots.

immunoblot assays (Reader Service No. 109). The range includes 1-, 5-, 10-, 15- and 20-nm conjugates of antibodies to mouse, human, rabbit and rat immunoglobulins, as well as Protein A and G. Silver-enhancing kits for light microscopy and immunoblotting assays are available. In addition, Caltag offers immunogold-conjugated F(ab')₂ fragments of goat anti-mouse IgG (H+L)-human adsorbed and goat anti-rabbit IgG (H+L)-human adsorbed, which, the company says, provide improved sensitivity and lower background than immunogold conjugates made from intact IgG antibodies. As both antibodies can be supplied conjugated with 5- and 10-nm particles, they are suitable for electron microscopy and light microscopy.

To suit individual needs, Janssen Biochimica is offering a range of new Omni-Stain kits for the **detection of primary antibodies** (Reader Service No. 110). The kits contain a highly reactive biotinylated link antibody and an enzyme-conjugated detection antibody. Researchers can choose between species-specific kits for goat, mouse and rabbit and between horseradish peroxidase and alkaline phosphatase conjugates. To provide high sensitivity and low background, the antibodies contain background-reducing stabilizing buffers and blocking agents, says Janssen. Each kit contains sufficient

ADVERTISEMENT

Teflon^{*} Coated and Bare Wire from

MEDWIRE

Medwire provides the highest quality platinum, platinum alloy, silver and stainless steel wire in both single and multistrand configurations.

For prompt delivery or additional information, please contact:

MEDWIRE, an affiliate of Sigmund Cohn Corp.
121 So. Columbus Ave. Mount Vernon, NY 10553
Tel: (914) 664-5300 Fax: (914) 664-5377

*Reg. TM of E.I. DuPont & Co.



ADVERTISEMENTS

Synthetic Oligonucleotides

Custom made Probes and Primers,
Synthesised, purified and shipped in 3 days

£4.00 per base

(All inclusive)
0.2 µM scale (approx. 500 µg)

Molecular Medicine Unit
King's College School of Medicine & Dentistry
Denmark Hill, London SE5 8RX

Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No.56

SPECIAL PRICE CUSTOM DNA

100 µg, 58.00 FF A BASE
FOR THE FIRST 15 BASES,
49.00 FF, each additional base
BIOPROBE SYSTEMS

CALL : 33-1-48-51-66-22

FAX : 33-1-48-51-59-90

Reader Service No.55

Custom DNA

Purified and Shipped in 48 hours,
Worldwide.

100 µg, \$10.00 a base for the first 15 bases,
\$7.50 for each additional base.
Delivered.

Research Genetics

1-800-533-4363

In the United Kingdom 0-800-89-1393

In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.26

DIABORM EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify
the binding parameters and
thermodynamic data between
ligands and biopolymers.
Results come closest to
the actual values.
Sample volume from
0.2 ml to 5.0 ml.
Up to 20 experiments
in parallel.

DIABORM
Gentle
P.O. Box 550126
D-8000 Munich 65,
FRG.

Reader Service No.49

ready-to-use reagents for 500 tests. Tests can be performed on formalin-fixed, paraffin-embedded or frozen sections, smears and cytospin preparations.

Staining solutions

Pro-Green is the new **reversible stain** for PAGE gels from Integrated Separation Systems (*Reader Service No. 111*). Using Pro-Green, researchers can check gel staining patterns before proceeding with western blotting or electroelution and without loss of resolution, says ISS. Pro-Green generates a negative image in five minutes without the need for a fixation



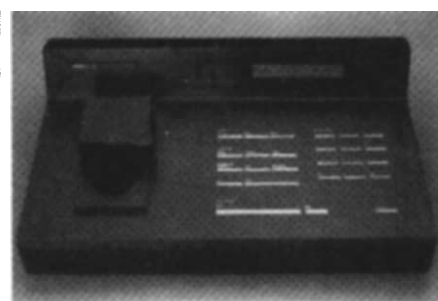
Pro-Green reversible stain for a quick check of gel banding patterns.

step. As the proteins are not treated with fixative, the process is fully reversible. Gels can be destained in five minutes. Pro-Green kits, which cost \$80 (US), can process about 20 gels.

Fluoro-Hance, is Research Products International's fast-acting **fluorographic reagent for enhancing autoradiographs** (*Reader Service No. 112*). According to RPI, Fluoro-Hance produces high-quality images from low-level beta-emitters such as ^3H , ^{14}C and ^{35}S . Costing \$36 (US) for a 500-ml quantity, Fluoro-Hance contains a high-efficiency scintillator system that is specially formulated for the rapid penetration of polyacrylamide and agarose gels. Fluoro-Hance offers one-step processing of gels and results in 30 minutes or less, says RPI.

Lab and safety equipment

Quick-Count from Bioscan provides a rapid method for **monitoring DNA labeling** without the need for scintillation fluid (*Reader Service No. 113*). The \$2,475 (US) microprocessor-controlled benchtop counter is a useful tool for molecular biologists as it can be used to count the radioisotopes ^{32}P , ^{125}I , ^{35}S and ^{14}C without the need for pipetting or sample preparation. A calibration procedure automatically converts c.p.m. to d.p.m. Quantitation is direct, with an accurate linear response to



Quick-Count for scintillation fluid-free counting of ^{125}I , ^{32}P , ^{14}C and ^{35}S .

5×10^9 d.p.m., says Bioscan. By using Quick-Count in conjunction with the company's flow analysis module, output from chromatography columns can be counted directly with a stated efficiency of 40 per cent. By viewing the elution profile in this way, the user has immediate access to information on the separation quality and percentage of ^{32}P incorporation.

Iso-Medic 4/600 HE is ICN's high-energy, multi-detector **gamma counter**, which is capable of simultaneously evaluating two isotopes having energies up to 1,500 KeV (*Reader Service No. 114*). The instrument's live-time electronics provide accurate and efficient counting of high-activity samples and provide a count linearity of one per cent or better for counting samples up to 4,000,000 d.p.m. When counting dual-label samples, the instrument automatically corrects for spilldown and crossover. Iso-Medic has a rack-advance system that can accommodate up to 600 tubes. Lead shielding positioned in front of and below the crystal detector arrays is designed to reduce background interference from high-energy samples on the deck. The instrument includes an on-board IBM 80286-compatible computer with 30-MByte hard-disk drive, 3.5-inch microfloppy disk drive, nine-inch colour monitor and slide-out keyboard.

Scotlab has added several new products to its range of **iodine radiation shielding devices** (*Reader Service No. 115*). The £215 (UK) protection shield is fabricated from 12-mm lead-impregnated acrylic and has a lead equivalence value of 0.5 mm. This is sufficient to block ^{125}I radioactive emissions, says Scotlab. The ^{125}I shield is designed with a 15° tilt to provide body and eye protection to researchers sitting at work-bench level. Scotlab's £225 (UK) ^{125}I storage box shields the user against gamma-emitting isotopes up to the energy levels of ^{125}I . The iodine box can accommodate up to three of Scotlab's interchangeable work racks. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain additional information about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.